

ABC's of M-estimation



Paul Zivich,¹ Rachael Ross,² Bonnie Shook-Sa¹

¹University of North Carolina, ²Columbia University

Acknowledgements

I am a postdoctoral research scientist at Columbia University

Supported by NIH R01DA056407

✉ rachael.k.ross@columbia.edu

🐙 rachael-k-ross

🐦 @rachael_k_ross

Section 1: introduction

Break (15min)

Section 2: applied examples

Break (15min)

Section 3: in context

Overview: Section 2 - Applied Examples

1. Logistic regression
2. Standardization
 - Outcome model based standardization
 - Inverse probability weighting
3. Transporting a risk
 - Outcome model based standardization
 - Inverse odds weighting

Each section will include time to work with code

As the goal is to illustrate implementation, we will not dwell on identification assumptions of these examples¹

¹Hernan & Robins Causal Inference: What If; Cole et al. *AJE* 2012

Overview: Section 2 - Applied Examples

1. **Logistic regression**
2. Standardization
 - **Outcome model based standardization**
 - Inverse probability weighting
3. Transporting a risk
 - Outcome model based standardization
 - Inverse odds weighting

These examples have been published in an IJE Education Corner²

²Ross et al. M-estimation for common epidemiological measures: introduction and applied examples *IJE*, 2024, 53(2):1-6.

O_i : observed data for unit i

- $O_i = (W_i, X_i, Y_i)$
- $O_i = (S_i, W_i, X_i, S_i \times Y_i)$

$$\text{expit}(a) = 1/(1 + \exp(-a))$$

Estimating function

$$\psi(O_i; \theta)$$

Estimating equation

$$\sum_{i=1}^n \psi(O_i; \theta) = 0$$

Logistic regression

Example

Data from the Zambia Preterm Birth Prevention Study (ZAPPS)³

$O_i = (W_i, X_i, Y_i)$

W_i , elevated blood pressure

X_i , anemia

Y_i , preterm birth

We want to estimate the parameters of this model

$$\begin{aligned}\Pr(Y_i = 1|X_i, W_i) &= \text{expit}(\beta_0 + \beta_1 X_i + \beta_2 W_i) \\ &= \text{expit}(\mathbf{Z}_i \boldsymbol{\beta}^T)\end{aligned}$$

where $\mathbf{Z}_i = (1, X_i, W_i)$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$

³Castillo et al. *Gates Open Research* 2019

$n = 826$ pregnant people

14.2% delivered preterm

14.3% diagnosed with early-pregnancy anemia

18.9% diagnosed with early-pregnancy elevated blood pressure

Estimating β by MLE

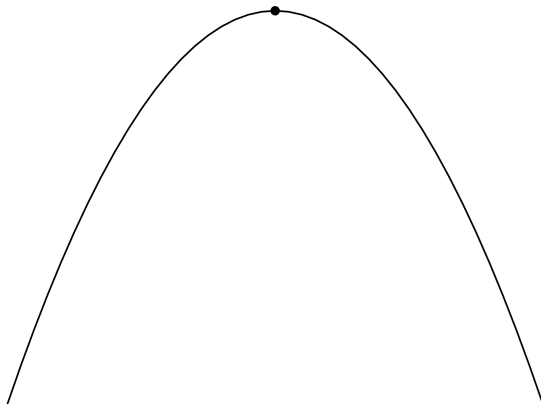
Let $\hat{\beta}_{mle}$ be the estimate of β by MLE

$\hat{\beta}_{mle}$ are the values that maximize the log-likelihood

$$\begin{aligned} & \sum_{i=1}^n \ln(L(\mathbf{O}_i, \hat{\beta}_{mle})) \\ &= \sum_{i=1}^n \left\{ Y_i \ln(\text{expit}(\mathbf{Z}_i \hat{\beta}_{mle}^T)) + (1 - Y_i) \ln(1 - \text{expit}(\mathbf{Z}_i \hat{\beta}_{mle}^T)) \right\} \end{aligned}$$

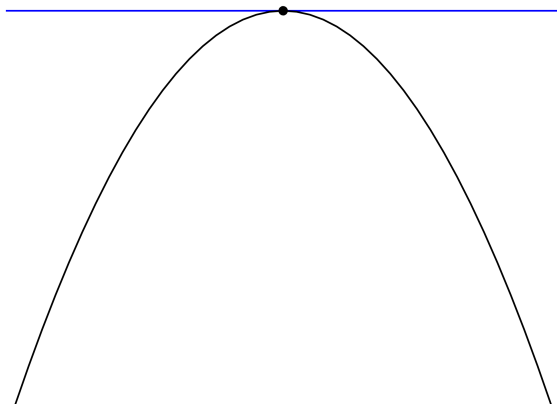
Estimating β by root-finding

Maximum of likelihood is where slope (derivative) is zero ("root")



Estimating β by root-finding

Maximum of likelihood is where slope (derivative) is zero ("root")



Estimating β by root-finding

Our log-likelihood is

$$\ln(L(\mathbf{O}_i, \beta)) = Y_i \ln(\text{expit}(\mathbf{Z}_i \beta^T)) + (1 - Y_i) \ln(1 - \text{expit}(\mathbf{Z}_i \beta^T))$$

Functions for the slope (derivative⁴)

$$\psi(\mathbf{O}_i, \beta) = \frac{\partial \ln(L(\mathbf{O}_i, \beta))}{\partial \beta} = \begin{bmatrix} (Y_i - \text{expit}(\mathbf{Z}_i \beta^T)) 1 \\ (Y_i - \text{expit}(\mathbf{Z}_i \beta^T)) X_i \\ (Y_i - \text{expit}(\mathbf{Z}_i \beta^T)) W_i \end{bmatrix}$$

⁴first-order partial derivatives, called the "score" functions

Estimating β by root-finding

Let $\hat{\beta}_{me}$ be the estimate of β by root finding (M-estimation)

Our estimating equation

$$\sum_{i=1}^n \psi(\mathbf{O}_i, \hat{\beta}_{me}) = \sum_{i=1}^n \begin{bmatrix} Y_i - \text{expit}(\mathbf{Z}_i \hat{\beta}_{me}^T) \\ (Y_i - \text{expit}(\mathbf{Z}_i \hat{\beta}_{me}^T)) X_i \\ (Y_i - \text{expit}(\mathbf{Z}_i \hat{\beta}_{me}^T)) W_i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

To implement, we need to define the estimating functions

Under **Defining estimating equation**

- R

```
p <- plogis(beta[1] + beta[2]*dat$anemia + beta[3]*dat$bp)

ef_1 <- (dat$ptb - p)
ef_2 <- (dat$ptb - p)*dat$anemia
ef_3 <- (dat$ptb - p)*dat$bp
```


$\hat{\beta}_{mle}$ and $\hat{\beta}_{me}$ from R

	β_0	β_1	β_2
$\hat{\beta}_{mle}$	-1.89450081814 3	0.1187353484 28	0.3605113262 76
$\hat{\beta}_{me}$	-1.89450081814 8	0.1187353484 32	0.3605113262 83

Difference is result of algorithm/tolerance

Let $\mathbf{V}(\hat{\beta})$ be the 3×3 covariance matrix of $\hat{\beta}$

$$\mathbf{V}(\hat{\beta}) = \begin{bmatrix} \text{var}(\hat{\beta}_0) & \text{cov}(\hat{\beta}_0, \hat{\beta}_1) & \text{cov}(\hat{\beta}_0, \hat{\beta}_2) \\ \text{cov}(\hat{\beta}_0, \hat{\beta}_1) & \text{var}(\hat{\beta}_1) & \text{cov}(\hat{\beta}_1, \hat{\beta}_2) \\ \text{cov}(\hat{\beta}_0, \hat{\beta}_2) & \text{cov}(\hat{\beta}_1, \hat{\beta}_2) & \text{var}(\hat{\beta}_2) \end{bmatrix}$$

The diagonal vector of \mathbf{V}/n (i.e., $\text{diag}(\mathbf{V}/n)$) are the finite sample variances

For Wald-type confidence intervals, we use $\text{SE}(\hat{\beta}) = \sqrt{\text{diag}(\mathbf{V}/n)}$

Estimating $V(\hat{\beta}_{mle})$ by MLE

$$V(\hat{\beta}_{mle}) = I(\hat{\beta}_{mle})^{-1}$$

where $I(\hat{\beta}_{mle})$ is the observed information matrix at $\hat{\beta}_{mle}$

Estimating $V(\hat{\beta}_{mle})$

There are two ways to estimate $I(\hat{\beta}_{mle})$

1. Hessian-based estimator

$$I(\hat{\beta}_{mle}) = \frac{1}{n} \sum_{i=1}^n \left[-\psi'(O_i, \hat{\beta}_{mle}) \right]$$

$$\text{where } \psi' = \frac{\partial \psi(O_i, \hat{\beta}_{mle})}{\partial \hat{\beta}_{mle}} = \frac{\partial^2 \ln(L(O_i, \hat{\beta}_{mle}))}{\partial \hat{\beta}_{mle}^2}$$

There are two ways to estimate $I(\hat{\beta}_{mle})$

2. Residual-based estimator

$$I(\hat{\beta}_{mle}) = \frac{1}{n} \sum_{i=1}^n \left[\psi(\mathbf{O}_i, \hat{\beta}_{mle}) \psi(\mathbf{O}_i, \hat{\beta}_{mle})^T \right]$$

Estimating $V(\hat{\beta}_{mle})$

There are two ways to estimate $I(\hat{\beta}_{mle})$

1. Hessian-based estimator

$$I(\hat{\beta}_{mle}) = \frac{1}{n} \sum_{i=1}^n \left[-\psi'(\mathbf{O}_i, \hat{\beta}_{mle}) \right]$$

2. Residual-based estimator

$$I(\hat{\beta}_{mle}) = \frac{1}{n} \sum_{i=1}^n \left[\psi(\mathbf{O}_i, \hat{\beta}_{mle}) \psi(\mathbf{O}_i, \hat{\beta}_{mle})^T \right]$$

Asymptotically equal when chosen parametric family is correct
Hessian is more efficient and default (logistic, glm, GLM)

Estimating $V(\hat{\beta}_{me})$ by sandwich variance estimator

$$V(\hat{\beta}_{me}) = B(\hat{\beta}_{me})^{-1} F(\hat{\beta}_{me}) [B(\hat{\beta}_{me})^{-1}]^T$$

Recall

- $B(\hat{\beta}) = \frac{1}{n} \sum_i \left[-\psi'(O_i, \hat{\beta}) \right]$

This corresponds to the Hessian-based info matrix estimator

- $F(\hat{\beta}) = \frac{1}{n} \sum_i \left[\psi(O_i, \hat{\beta}) \psi(O_i, \hat{\beta})^T \right]$

This corresponds to the residual-based info matrix estimator

Asymptotic equivalence

When chosen parametric family is correct

$$\begin{aligned} \mathbf{V}(\hat{\beta}) &= \mathbf{B}(\hat{\beta})^{-1} \mathbf{F}(\hat{\beta}) [\mathbf{B}(\hat{\beta})^{-1}]^T \\ &\sim \mathbf{I}(\hat{\beta})^{-1} \mathbf{I}(\hat{\beta}) [\mathbf{I}(\hat{\beta})^{-1}]^T \\ &= \mathbf{I}(\hat{\beta})^{-1} \end{aligned}$$

Estimating $V(\hat{\beta})$

1. Use the same code for the estimating equation
2. Baking the bread
 - Use software to estimate the Hessian
3. Cooking the filling
 - Use transpose and dot product functions
4. Assembling the sandwich
 - Use inverse function

Standard errors from R

	β_0	β_1	β_2
MLE, Hessian-based	0.01 4 96046	0.077 6 4837	0.056 6 0453
MLE, Residual-based	0.01 5 08328	0.077 6 1366	0.056 7 0628
Sandwich estimator	0.01 4 84043	0.077 7 2035	0.056 5 2969

Time to work with code

Standardization

Data and estimand

Using the same data as the prior example

$$O_i = (Y_i, X_i, W_i)$$

Y_i , preterm birth

X_i , anemia

W_i , elevated blood pressure

We want to estimate the marginal risk difference and risk ratio of anemia on preterm birth, standardized by elevated blood pressure.

$$E(E(Y|X = 1, W)|W) - E(E(Y|X = 0, W)|W)$$

$$E(E(Y|X = 1, W)|W)/E(E(Y|X = 0, W)|W)$$

We will illustrate two standardization approaches

- Outcome model based standardization (i.e., g-computation)
- Inverse probability weighting

Outcome model based standardization

1. Estimate outcome model (from prior example)

$$\Pr(Y_i = 1|X_i, W_i) = \text{expit}(\beta_0 + \beta_1 X_i + \beta_2 W_i) = \text{expit}(\mathbf{Z}_i \boldsymbol{\beta}^T)$$

Outcome model based standardization

1. Estimate outcome model (from prior example)

$$\Pr(Y_i = 1|X_i, W_i) = \text{expit}(\beta_0 + \beta_1 X_i + \beta_2 W_i) = \text{expit}(\mathbf{Z}_i \boldsymbol{\beta}^T)$$

2. Use $\hat{\boldsymbol{\beta}}$ to predict outcome when $X_i = 1$. Take the mean.

$$\hat{\mu}_1 = \frac{1}{n} \sum_{i=1}^n \text{expit}(\hat{\beta}_0 + \hat{\beta}_1 \times 1 + \hat{\beta}_2 W_i)$$

Outcome model based standardization

1. Estimate outcome model (from prior example)

$$\Pr(Y_i = 1|X_i, W_i) = \text{expit}(\beta_0 + \beta_1 X_i + \beta_2 W_i) = \text{expit}(\mathbf{Z}_i \boldsymbol{\beta}^T)$$

2. Use $\hat{\boldsymbol{\beta}}$ to predict outcome when $X_i = 1$. Take the mean.

$$\hat{\mu}_1 = \frac{1}{n} \sum_{i=1}^n \text{expit}(\hat{\beta}_0 + \hat{\beta}_1 \times 1 + \hat{\beta}_2 W_i)$$

3. Use $\hat{\boldsymbol{\beta}}$ to predict outcome when $X_i = 0$. Take the mean.

$$\hat{\mu}_2 = \frac{1}{n} \sum_{i=1}^n \text{expit}(\hat{\beta}_0 + \hat{\beta}_1 \times 0 + \hat{\beta}_2 W_i)$$

Outcome model based standardization

1. Estimate outcome model (from prior example)

$$\Pr(Y_i = 1|X_i, W_i) = \text{expit}(\beta_0 + \beta_1 X_i + \beta_2 W_i) = \text{expit}(\mathbf{Z}_i \boldsymbol{\beta}^T)$$

2. Use $\hat{\boldsymbol{\beta}}$ to predict outcome when $X_i = 1$. Take the mean.

$$\hat{\mu}_1 = \frac{1}{n} \sum_{i=1}^n \text{expit}(\hat{\beta}_0 + \hat{\beta}_1 \times 1 + \hat{\beta}_2 W_i)$$

3. Use $\hat{\boldsymbol{\beta}}$ to predict outcome when $X_i = 0$. Take the mean.

$$\hat{\mu}_2 = \frac{1}{n} \sum_{i=1}^n \text{expit}(\hat{\beta}_0 + \hat{\beta}_1 \times 0 + \hat{\beta}_2 W_i)$$

4. Calculate the risk difference $\hat{\delta}_1 = \hat{\mu}_1 - \hat{\mu}_2$ & log risk ratio $\hat{\delta}_2 = \ln(\hat{\mu}_1/\hat{\mu}_2)$

Outcome model based standardization

Our stacked estimating equation

$$\sum_{i=1}^n \psi(\mathbf{O}_i, \hat{\boldsymbol{\theta}}) = \sum_{i=1}^n \begin{bmatrix} Y_i - \text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\beta}}^T) \\ (Y_i - \text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\beta}}^T)) X_i \\ (Y_i - \text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\beta}}^T)) W_i \\ \text{expit}(\hat{\beta}_0 + \hat{\beta}_1 \times 1 + \hat{\beta}_2 W_i) - \hat{\mu}_1 \\ \text{expit}(\hat{\beta}_0 + \hat{\beta}_1 \times 0 + \hat{\beta}_2 W_i) - \hat{\mu}_2 \\ (\hat{\mu}_1 - \hat{\mu}_2) - \hat{\delta}_1 \\ \ln(\hat{\mu}_1 / \hat{\mu}_2) - \hat{\delta}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\delta})$

Outcome model based standardization

Our stacked estimating equation

$$\sum_{i=1}^n \psi(\mathbf{O}_i, \hat{\boldsymbol{\theta}}) = \sum_{i=1}^n \begin{bmatrix} Y_i - \text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\beta}}^T) \\ (Y_i - \text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\beta}}^T)) X_i \\ (Y_i - \text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\beta}}^T)) W_i \\ \text{expit}(\hat{\beta}_0 + \hat{\beta}_1 \times 1 + \hat{\beta}_2 W_i) - \hat{\mu}_1 \\ \text{expit}(\hat{\beta}_0 + \hat{\beta}_1 \times 0 + \hat{\beta}_2 W_i) - \hat{\mu}_2 \\ (\hat{\mu}_1 - \hat{\mu}_2) - \hat{\delta}_1 \\ \ln(\hat{\mu}_1 / \hat{\mu}_2) - \hat{\delta}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\delta})$

Outcome model based standardization

Our stacked estimating equation

$$\sum_{i=1}^n \psi(\mathbf{O}_i, \hat{\boldsymbol{\theta}}) = \sum_{i=1}^n \begin{bmatrix} Y_i - \text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\beta}}^T) \\ (Y_i - \text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\beta}}^T)) X_i \\ (Y_i - \text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\beta}}^T)) W_i \\ \text{expit}(\hat{\beta}_0 + \hat{\beta}_1 \times 1 + \hat{\beta}_2 W_i) - \hat{\mu}_1 \\ \text{expit}(\hat{\beta}_0 + \hat{\beta}_1 \times 0 + \hat{\beta}_2 W_i) - \hat{\mu}_2 \\ (\hat{\mu}_1 - \hat{\mu}_2) - \hat{\delta}_1 \\ \ln(\hat{\mu}_1 / \hat{\mu}_2) - \hat{\delta}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\delta})$

Outcome model based standardization

Our stacked estimating equation

$$\sum_{i=1}^n \psi(\mathbf{O}_i, \hat{\boldsymbol{\theta}}) = \sum_{i=1}^n \begin{bmatrix} Y_i - \text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\beta}}^T) \\ (Y_i - \text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\beta}}^T)) X_i \\ (Y_i - \text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\beta}}^T)) W_i \\ \text{expit}(\hat{\beta}_0 + \hat{\beta}_1 \times 1 + \hat{\beta}_2 W_i) - \hat{\mu}_1 \\ \text{expit}(\hat{\beta}_0 + \hat{\beta}_1 \times 0 + \hat{\beta}_2 W_i) - \hat{\mu}_2 \\ (\hat{\mu}_1 - \hat{\mu}_2) - \hat{\delta}_1 \\ \ln(\hat{\mu}_1 / \hat{\mu}_2) - \hat{\delta}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\delta})$

Outcome model based standardization

Our stacked estimating equation

$$\sum_{i=1}^n \psi(\mathbf{O}_i, \hat{\boldsymbol{\theta}}) = \sum_{i=1}^n \begin{bmatrix} Y_i - \text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\beta}}^T) \\ (Y_i - \text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\beta}}^T)) X_i \\ (Y_i - \text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\beta}}^T)) W_i \\ \text{expit}(\hat{\beta}_0 + \hat{\beta}_1 \times 1 + \hat{\beta}_2 W_i) - \hat{\mu}_1 \\ \text{expit}(\hat{\beta}_0 + \hat{\beta}_1 \times 0 + \hat{\beta}_2 W_i) - \hat{\mu}_2 \\ (\hat{\mu}_1 - \hat{\mu}_2) - \hat{\delta}_1 \\ \ln(\hat{\mu}_1 / \hat{\mu}_2) - \hat{\delta}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\delta})$

Note: we estimate the log risk ratio because 1) it does not have bounds which is better for root-finding and 2) we want to estimate the variance of the log risk ratio for constructing confidence intervals

Outcome model based standardization

Our stacked estimating equation

$$\sum_{i=1}^n \psi(\mathbf{O}_i, \hat{\boldsymbol{\theta}}) = \sum_{i=1}^n \begin{bmatrix} Y_i - \text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\beta}}^T) \\ (Y_i - \text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\beta}}^T)) X_i \\ (Y_i - \text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\beta}}^T)) W_i \\ \text{expit}(\hat{\beta}_0 + \hat{\beta}_1 \times 1 + \hat{\beta}_2 W_i) - \hat{\mu}_1 \\ \text{expit}(\hat{\beta}_0 + \hat{\beta}_1 \times 0 + \hat{\beta}_2 W_i) - \hat{\mu}_2 \\ (\hat{\mu}_1 - \hat{\mu}_2) - \hat{\delta}_1 \\ \ln(\hat{\mu}_1 / \hat{\mu}_2) - \hat{\delta}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\delta})$

Estimating functions: Code

- R

```
p <- plogis(beta[1] + beta[2]*dat$anemia + beta[3]*dat$bp)
ef_1 <- (dat$ptb - p)
ef_2 <- (dat$ptb - p)*dat$anemia
ef_3 <- (dat$ptb - p)*dat$bp

ef_r1 <- plogis(beta[1] + beta[2]*1 + beta[3]*dat$bp) - mu[1]
ef_r0 <- plogis(beta[1] + beta[2]*0 + beta[3]*dat$bp) - mu[2]

ef_rd <- (mu[1] - mu[2]) - delta[1]
ef_lnr <- log(mu[1]/mu[2]) - delta[2]
```


Inverse probability weighting

1. Estimate propensity score model

$$\Pr(X_i = 1|W_i) = \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)$$

Inverse probability weighting

1. Estimate propensity score model

$$\Pr(X_i = 1|W_i) = \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)$$

2. Among those with $X_i = 1$, take a mean of the outcome, weighted by the inverse probability of $X_i = 1$,

$$\hat{\mu}_1 = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{X_i Y_i}{\text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)} \right\}$$

Note: This is the Horvitz-Thompson estimator

Inverse probability weighting

1. Estimate propensity score model

$$\Pr(X_i = 1|W_i) = \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)$$

2. Among those with $X_i = 1$, take a mean of the outcome, weighted by the inverse probability of $X_i = 1$,

$$\hat{\mu}_1 = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{X_i Y_i}{\text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)} \right\}$$

3. Among those with $X_i = 0$, take a mean of the outcome, weighted by the inverse probability of $X_i = 0$,

$$\hat{\mu}_2 = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{(1 - X_i) Y_i}{1 - \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)} \right\}$$

Inverse probability weighting

1. Estimate propensity score model

$$\Pr(X_i = 1|W_i) = \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)$$

2. Among those with $X_i = 1$, take a mean of the outcome, weighted by the inverse probability of $X_i = 1$,

$$\hat{\mu}_1 = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{X_i Y_i}{\text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)} \right\}$$

3. Among those with $X_i = 0$, take a mean of the outcome, weighted by the inverse probability of $X_i = 0$,

$$\hat{\mu}_2 = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{(1 - X_i) Y_i}{1 - \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)} \right\}$$

4. Calculate the risk difference $\hat{\delta}_1 = \hat{\mu}_1 - \hat{\mu}_2$ & log risk ratio $\hat{\delta}_2 = \ln(\hat{\mu}_1/\hat{\mu}_2)$

Inverse probability weighting

Our stacked estimating equation

$$\sum_{i=1}^n \psi(O_i, \hat{\theta}) = \sum_{i=1}^n \begin{bmatrix} X_i - \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i) \\ (X_i - \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)) W_i \\ \frac{X_i Y_i}{\text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)} - \hat{\mu}_1 \\ \frac{(1-X_i) Y_i}{1 - \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)} - \hat{\mu}_2 \\ (\hat{\mu}_1 - \hat{\mu}_2) - \hat{\delta}_1 \\ \ln(\hat{\mu}_1 / \hat{\mu}_2) - \hat{\delta}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

where $\theta = (\alpha, \mu, \delta)$

Inverse probability weighting

Our stacked estimating equation

$$\sum_{i=1}^n \psi(O_i, \hat{\theta}) = \sum_{i=1}^n \begin{bmatrix} X_i - \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i) \\ (X_i - \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)) W_i \\ \frac{X_i Y_i}{\text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)} - \hat{\mu}_1 \\ \frac{(1-X_i) Y_i}{1 - \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)} - \hat{\mu}_2 \\ (\hat{\mu}_1 - \hat{\mu}_2) - \hat{\delta}_1 \\ \ln(\hat{\mu}_1 / \hat{\mu}_2) - \hat{\delta}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

where $\theta = (\alpha, \mu, \delta)$

Inverse probability weighting

Our stacked estimating equation

$$\sum_{i=1}^n \psi(O_i, \hat{\theta}) = \sum_{i=1}^n \begin{bmatrix} X_i - \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i) \\ (X_i - \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)) W_i \\ \frac{X_i Y_i}{\text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)} - \hat{\mu}_1 \\ \frac{(1 - X_i) Y_i}{1 - \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)} - \hat{\mu}_2 \\ (\hat{\mu}_1 - \hat{\mu}_2) - \hat{\delta}_1 \\ \ln(\hat{\mu}_1 / \hat{\mu}_2) - \hat{\delta}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

where $\theta = (\alpha, \mu, \delta)$

Inverse probability weighting

Our stacked estimating equation

$$\sum_{i=1}^n \psi(O_i, \hat{\theta}) = \sum_{i=1}^n \begin{bmatrix} X_i - \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i) \\ (X_i - \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)) W_i \\ \frac{X_i Y_i}{\text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)} - \hat{\mu}_1 \\ \frac{(1-X_i) Y_i}{1 - \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)} - \hat{\mu}_2 \\ (\hat{\mu}_1 - \hat{\mu}_2) - \hat{\delta}_1 \\ \ln(\hat{\mu}_1 / \hat{\mu}_2) - \hat{\delta}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

where $\theta = (\alpha, \mu, \delta)$

Inverse probability weighting

Our stacked estimating equation

$$\sum_{i=1}^n \psi(O_i, \hat{\theta}) = \sum_{i=1}^n \begin{bmatrix} X_i - \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i) \\ (X_i - \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)) W_i \\ \frac{X_i Y_i}{\text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)} - \hat{\mu}_1 \\ \frac{(1-X_i) Y_i}{1 - \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)} - \hat{\mu}_2 \\ (\hat{\mu}_1 - \hat{\mu}_2) - \hat{\delta}_1 \\ \ln(\hat{\mu}_1 / \hat{\mu}_2) - \hat{\delta}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

where $\theta = (\alpha, \mu, \delta)$

Estimating functions: Code

- R

```
pscore <- plogis(alpha[1] + alpha[2]*bp)
ef_1 <- (anemia - pscore)
ef_2 <- (anemia - pscore)*bp

wt <- anemia/pscore + (1-anemia)/(1-pscore)
ef_r1 <- anemia*wt*ptb - mu[1]
ef_r0 <- (1 - anemia)*wt*ptb - mu[2]

ef_rd <- (mu[1] - mu[2]) - delta[1]
ef_lnr <- log(mu[1]/mu[2]) - delta[2]
```

Aside: IPW GEE trick for conservative SE/intervals

The propensity score parameters are excluded from the stack of estimating functions, i.e. treats α parameters as known

$$\sum_{i=1}^n \psi(O_i, \hat{\theta}) = \sum_{i=1}^n \begin{bmatrix} \cancel{X_i - \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)} \\ \cancel{(X_i - \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)) W_i} \\ \frac{X_i Y_i}{\text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)} - \hat{\mu}_1 \\ \frac{(1-X_i) Y_i}{1 - \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 W_i)} - \hat{\mu}_2 \\ (\hat{\mu}_1 - \hat{\mu}_2) - \hat{\delta}_1 \\ \ln(\hat{\mu}_1 / \hat{\mu}_2) - \hat{\delta}_2 \end{bmatrix}$$

Then use sandwich variance estimator⁵

⁵Robins et al. Marginal Structural Models and Causal Inference in Epidemiology *Epidemiology* 2000.

Time to work with code

Transporting a risk

Using the same data as the prior examples (ZAPPS) PLUS external target sample (simulated)

$$O_i = (S_i, X_i, W_i, S_i \times Y_i)$$

S_i , sample indicator: $S_i = 1$ for ZAPPS, $S_i = 0$ for external target

X_i , anemia

W_i , elevated blood pressure

Y_i , preterm birth, which is only observed when $S_i = 1$

We want to estimate the risk of preterm birth in the external target sample, i.e., $\Pr(Y = 1|S = 0)$.

We will illustrate two (standardization) approaches

- Outcome model based standardization
- Inverse odds weighting

For simplicity, we assume early-pregnancy anemia and blood pressure are sufficient to identify this transported risk

Outcome model based standardization

- 1 Estimate outcome model **in the ZAPPS data** (as in prior examples)

$$\begin{aligned}\Pr(Y_i = 1|X_i, W_i, S_i = 1) &= \text{expit}(\hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 W_i) \\ &= \text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\beta}}^T)\end{aligned}$$

- 2 Use $\hat{\boldsymbol{\beta}}$ to predict outcome **in the external target**. Take the mean.

$$\hat{\mu} = \frac{1}{\sum_{i=1}^n (1 - S_i)} \sum_{i=1}^n (1 - S_i) \text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\beta}}^T)$$

Our estimating equation

$$\sum_{i=1}^n \psi(\mathbf{O}_i, \hat{\boldsymbol{\theta}}) = \sum_{i=1}^n \begin{bmatrix} S_i(Y_i - \text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\beta}}^T)) \\ S_i(Y_i - \text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\beta}}^T))X_i \\ S_i(Y_i - \text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\beta}}^T))W_i \\ (1 - S_i)(\text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\beta}}^T) - \hat{\mu}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mu)$

Estimating functions: Code

- R

```
p <- plogis(beta[1] + beta[2]*anemia + beta[3]*bp)
```

```
ef_1 <- S*(ptb - p)
```

```
ef_2 <- S*(ptb - p)*anemia
```

```
ef_3 <- S*(ptb - p)*bp
```

```
ef_r <- (1-S)*(p - mu[1])
```

Inverse odds weighting

1. Estimate selection model

$$\begin{aligned}\Pr(S_i = 1 | X_i, W_i) &= \text{expit}(\hat{\gamma}_0 + \hat{\gamma}_1 X_i + \hat{\gamma}_2 W_i) \\ &= \text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\gamma}}^T)\end{aligned}$$

2. Among those with $S_i = 1$, take the weighted mean of the outcome, weighted by the inverse odds of $S_i = 1$

$$\hat{\mu} = \frac{1}{\sum_{i=1}^n S_i \pi_S(Z; \hat{\boldsymbol{\gamma}})} \sum_{i=1}^n S_i Y_i \pi_S(Z; \hat{\boldsymbol{\gamma}})$$

$$\text{where } \pi_S(Z; \hat{\boldsymbol{\gamma}}) = \frac{1 - \text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\gamma}}^T)}{\text{expit}(\mathbf{Z}_i \hat{\boldsymbol{\gamma}}^T)}$$

Note: This is a Hajek estimator. We used a Horvitz-Thompson estimator in the previous section

Our stacked estimating equation

$$\sum_{i=1}^n \psi(O_i, \hat{\theta}) = \sum_{i=1}^n \begin{bmatrix} S_i - \text{expit}(\mathbf{Z}_i \hat{\gamma}^T) \\ (S_i - \text{expit}(\mathbf{Z}_i \hat{\gamma}^T)) X_i \\ (S_i - \text{expit}(\mathbf{Z}_i \hat{\gamma}^T)) W_i \\ S_i \pi_S(Z; \hat{\gamma}) (Y_i - \mu) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

where $\theta = (\gamma, \mu)$ and $\pi_S(Z; \hat{\gamma}) = \frac{1 - \text{expit}(\mathbf{Z}_i \hat{\gamma}^T)}{\text{expit}(\mathbf{Z}_i \hat{\gamma}^T)}$

Estimating functions: Code

- R

```
p <- plogis(gamma[1] + gamma[2]*anemia + gamma[3]*bp)
ef_1 <- (S - p)
ef_2 <- (S - p)*anemia
ef_3 <- (S - p)*bp

oddswt <- (1-p)/p
ef_r <- S*oddswt*(ptb - mu[1])
```

Time to work with code

Section 1: introduction

Break (15min)

Section 2: applied examples

Break (15min)

Section 3: in context