Marginal causal effects

Arvid Sjölander

Department of Medical Epidemiology and Biostatistics
Karolinska Institutet

Causal Inference for Epidemiological Research

Exchangeability

- If we have exchangeability

$$(Y_0, Y_1) \perp\!\!\!\perp X,$$

  then the crude association is a causal effect
- In observational studies, we typically don't have exchangeability because of confounding
- If we have conditional exchangeability, given Z,

$$(Y_0, Y_1) \perp\!\!\!\perp X \mid Z,$$

  then controlling for Z gives a causal effect

Conditional effects

- When Z is low dimensional (e.g. binary or categorical with few levels), we may control for Z by stratification
  - conceptually simple
  - computationally simple
  - does not require any modeling assumptions
- Stratification gives one causal effect for each stratum - conditional (subpopulation) effects
  - e.g. stratification on sex gives the effect for men and women separately

Marginal effects

- Often, it may be desirable to estimate the causal effect for the whole population - a **marginal** causal effect
  - easier to interpret and communicate one marginal effect than several conditional effects
  - randomized trials give marginal effects, and we may want to make results from observational studies comparable
  - we may want to consider future interventions to the whole population, rather than to subgroups
- We will consider two methods for estimation of marginal effects
  - standardization
  - inverse probability weighting

Outline

Estimation without regression models
- Standardization
- Inverse probability weighting

Estimation with regression models
- Standardization
- Inverse probability weighting

Outline

Estimation without regression models
- Standardization
- Inverse probability weighting

Estimation with regression models
- Standardization
- Inverse probability weighting

Outline

Estimation without regression models
- Standardization
- Inverse probability weighting

Estimation with regression models
- Standardization
- Inverse probability weighting

Example (recapitulation)

	$Z = 0$		$Z = 1$	
	$Y = 0$	$Y = 1$	$Y = 0$	$Y = 1$
$X = 0$	270	30	120	80
$X = 1$	180	20	60	240

- *Assume conditional exchangeability, given Z, and compute the conditional causal risk ratio, given Z, for $Z = 1$ and $Z = 0$.*
- *Where in the computation do you use the assumption of exchangeability?*

Solution

	$Z=0$		$Z=1$	
	$Y=0$	$Y=1$	$Y=0$	$Y=1$
$X=0$	270	30	120	80
$X=1$	180	20	60	240

Conditional causal risk ratio, given $Z=0$:

$$\begin{aligned}\frac{p(Y_1=1|Z=0)}{p(Y_0=1|Z=0)} &= \{(Y_0,Y_1)\amalg X|Z\} \\ &= \frac{p(Y_1=1|X=1,Z=0)}{p(Y_0=1|X=0,Z=0)} = \frac{p(Y=1|X=1,Z=0)}{p(Y=1|X=0,Z=0)} \\ &= \frac{20/200}{30/300} = 1\end{aligned}$$

Solution

	$Z=0$		$Z=1$	
	$Y=0$	$Y=1$	$Y=0$	$Y=1$
$X=0$	270	30	120	80
$X=1$	180	20	60	240

Conditional causal risk ratio, given $Z=1$:

$$\begin{aligned}\frac{p(Y_1=1|Z=1)}{p(Y_0=1|Z=1)} &= \{(Y_0,Y_1)\amalg X|Z\} \\ &= \frac{p(Y_1=1|X=1,Z=1)}{p(Y_0=1|X=0,Z=1)} = \frac{p(Y=1|X=1,Z=1)}{p(Y=1|X=0,Z=1)} \\ &= \frac{240/300}{80/200} = 2\end{aligned}$$

Solution

- Given $Z=0$, the conditional causal risk ratio is equal to 1 (no effect)
- Given $Z=1$, the conditional causal risk ratio is equal to 2
- Effect modification by (interaction with) Z

The standardization formula

- If we have conditional exchangeability, given Z, then $p(Y_x=1)$ can be computed with the standardization formula

$$p(Y_x=1) = \sum_Z p(Y=1|X=x,Z)p(Z)$$

- Special case; binary Z:

$$\begin{aligned}p(Y_x=1) &= p(Y=1|X=x,Z=0)p(Z=0) \\ &+ p(Y=1|X=x,Z=1)p(Z=1)\end{aligned}$$

Proof

- Law of total probability:

$$p(Y_x = 1) = \sum_Z p(Y_x = 1|Z)p(Z)$$

- Conditional exchangeability, given Z:

$$\sum_Z p(Y_x = 1|Z)p(Z) = \sum_Z p(Y_x = 1|X = x, Z)p(Z)$$

- Definition of potential outcomes:

$$\sum_Z p(Y_x = 1|X = x, Z)p(Z) = \sum_Z p(Y = 1|X = x, Z)p(Z)$$

Example

$$p(Y_x = 1) = \sum_Z p(Y = 1|X = x, Z)p(Z)$$

	$Z = 0$		$Z = 1$	
	$Y = 0$	$Y = 1$	$Y = 0$	$Y = 1$
$X = 0$	270	30	120	80
$X = 1$	180	20	60	240

- *Assume conditional exchangeability, given Z, and compute the marginal causal risk ratio*

Solution

	$Z = 0$		$Z = 1$	
	$Y = 0$	$Y = 1$	$Y = 0$	$Y = 1$
$X = 0$	270	30	120	80
$X = 1$	180	20	60	240

$$p(Y_0 = 1) = \{(Y_0, Y_1) \amalg X|Z\} = \sum_Z p(Y = 1|X = 0, Z)p(Z)$$

$$= \underbrace{30/300}_{p(Y=1|X=0,Z=0)} \times \underbrace{500/1000}_{p(Z=0)} + \underbrace{80/200}_{p(Y=1|X=0,Z=1)} \times \underbrace{500/1000}_{p(Z=1)} = 0.25$$

$$p(Y_1 = 1) = \{(Y_0, Y_1) \amalg X|Z\} = \sum_Z p(Y = 1|X = 1, Z)p(Z)$$

$$= \underbrace{20/200}_{p(Y=1|X=1,Z=0)} \times \underbrace{500/1000}_{p(Z=0)} + \underbrace{240/300}_{p(Y=1|X=1,Z=1)} \times \underbrace{500/1000}_{p(Z=1)} = 0.45$$

$$\frac{p(Y_1 = 1)}{p(Y_0 = 1)} = \frac{0.45}{0.25} = 1.8$$

Marginal effect vs. conditional effects

- In our example we had that $p(Z = 1) = p(Z = 0) = 0.5$
- We observed that

$$\begin{aligned} \frac{p(Y_1 = 1|Z = 0)}{p(Y_0 = 1|Z = 0)} &= 1 \\ \frac{p(Y_1 = 1|Z = 1)}{p(Y_0 = 1|Z = 1)} &= 2 \\ \frac{p(Y_1 = 1)}{p(Y_0 = 1)} &= 1.8 \end{aligned}$$

- The marginal effect is not generally equal to the average of the conditional effects
  - which in our example would be $1 \times 0.5 + 2 \times 0.5 = 1.5$
- A special case; the marginal causal risk difference is equal to the average of the conditional causal risk differences

Technical note: non-collapsibility

- The marginal effect is not necessarily equal to average of the conditional effects - **even if these are constant across levels of** Z
  - e.g. the causal odds ratio may be equal to 3 for both men and women, but
  - the marginal causal odds ratio may be equal to 2
- This phenomenon is sometimes referred to as 'non-collapsibility'
- Odds ratios and hazard ratios are non-collapsible, whereas risk difference and risk ratios are not

Outline

Estimation without regression models
Standardization
Inverse probability weighting

Estimation with regression models
Standardization
Inverse probability weighting

An alternative method for marginal effects

- Inverse probability weighting (IPW) is an alternative method to compute the marginal causal effect
- Without modeling assumptions, IPW gives the same result as standardization
- IPW may give different results, and may sometimes be advantageous, when using regression models
  - more later

Three steps for IPW

- **Step 1**: for each level of the exposure X and confounders Z, compute the probability $p(X|Z)$
- **Step 2**: assign a weight to each subject i, equal to

$$W_i = \frac{1}{p(X_i|Z_i)}$$

  where X_i and Z_i are the observed exposure and confounders levels, respectively for subject i
  - for instance, suppose that $p(X = 1|Z = 1) = 0.2$
  - each subject with $(X = 1, Z = 1)$ is then counted as $1/0.2 = 5$ subjects in the analysis, and
  - each subjects with $(X = 0, Z = 1)$ is then counted as $1/(1 - 0.2) = 1.25$ subjects in the analysis
- **Step 3**: use $p(Y = 1|X = x)$ in the weighted sample as an estimate of $p(Y_x = 1)$

Example

	$Z=0$		$Z=1$	
	$Y=0$	$Y=1$	$Y=0$	$Y=1$
$X=0$	270	30	120	80
$X=1$	180	20	60	240

- ***Step 1**: for each level of the exposure X and confounders Z, compute the probability $p(X|Z)$*

Solution

	$Z=0$		$Z=1$	
	$Y=0$	$Y=1$	$Y=0$	$Y=1$
$X=0$	270	30	120	80
$X=1$	180	20	60	240

$$
\begin{aligned}
p(X=0|Z=0) &= 300/500 = 0.6 \\
p(X=1|Z=0) &= 200/500 = 0.4 \\
p(X=0|Z=1) &= 200/500 = 0.4 \\
p(X=1|Z=1) &= 300/500 = 0.6
\end{aligned}
$$

Example

	$Z=0$		$Z=1$	
	$Y=0$	$Y=1$	$Y=0$	$Y=1$
$X=0$	270	30	120	80
$X=1$	180	20	60	240

- ***Step 2**: assign a weight to each subject i, equal to*

$$W_i = \frac{1}{p(X_i|Z_i)}$$

Solution

	$Z=0$		$Z=1$	
	$Y=0$	$Y=1$	$Y=0$	$Y=1$
$X=0$	$\frac{270}{0.6}=450$	$\frac{30}{0.6}=50$	$\frac{120}{0.4}=300$	$\frac{80}{0.4}=200$
$X=1$	$\frac{180}{0.4}=450$	$\frac{20}{0.4}=50$	$\frac{60}{0.6}=100$	$\frac{240}{0.6}=400$

Example

	$Z=0$		$Z=1$	
	$Y=0$	$Y=1$	$Y=0$	$Y=1$
$X=0$	450	50	300	200
$X=1$	450	50	100	400

- ***Step 3**: use $p(Y=1|X=x)$ in the weighted sample as an estimate of $p(Y_x=1)$*

Solution

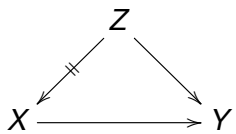
	$Z=0$		$Z=1$	
	$Y=0$	$Y=1$	$Y=0$	$Y=1$
$X=0$	450	50	300	200
$X=1$	450	50	100	400

$$
\begin{aligned}
p(Y_0=1) &= p(Y=1|X=0) = \frac{50+200}{450+50+300+200} = 0.25 \\
p(Y_1=1) &= p(Y=1|X=1) = \frac{50+400}{450+50+100+400} = 0.45 \\
\frac{p(Y_1=1)}{p(Y_0=1)} &= \frac{0.45}{0.25} = 1.8
\end{aligned}
$$

- Same as with standardization

Why IPW works

- IPW breaks the association between X and Z



- As a consequence, Z is not a confounder in the weighted sample
  - because a confounder must be associated with the exposure
- If Z is sufficient for confounding control in the original sample, then there is no confounding in the weighted sample
- Thus, in the weighted sample we have exchangeability, so that $p(Y=1|X=x) = p(Y_x=1)$

Example

	$Z=0$		$Z=1$	
	$Y=0$	$Y=1$	$Y=0$	$Y=1$
$X=0$	450	50	300	200
$X=1$	450	50	100	400

- *Verify that X and Z are independent in the weighted sample*

Solution

	$Z=0$		$Z=1$	
	$Y=0$	$Y=1$	$Y=0$	$Y=1$
$X=0$	450	50	300	200
$X=1$	450	50	100	400

$$p(X=1|Z=0) = \frac{450+50}{450+50+450+50} = 0.5$$
$$p(X=1|Z=1) = \frac{100+400}{300+200+100+400} = 0.5$$

- $p(X=1|Z=0) = p(X=1|Z=1)$ so X and Z are independent

Outline

Estimation without regression models
- Standardization
- Inverse probability weighting

Estimation with regression models
- Standardization
- Inverse probability weighting

Outline

Estimation without regression models
- Standardization
- Inverse probability weighting

Estimation with regression models
- Standardization
- Inverse probability weighting

Motivating example

- Suppose we carry out an observational study to estimate the causal effect of AZT on infection risk for AIDS patients
- 1000 subjects enrolled
- Baseline measures:
  - CD4 count (Z; counts/μl)
  - AZT level (X; '0' for 'untreated', '1' for 'treated')
- At end of follow up we measure:
  - infection status (Y; '1' for infection, '0' for no infection)

Data (in R)

```
> aids <- read.table("aids.txt", header=TRUE)
> aids[1:10, ]
     Z X Y
1  405 0 1
2  412 0 0
3  301 1 0
4  253 1 0
5  307 0 1
6  392 0 1
7  361 0 0
8  363 1 0
9  267 1 0
10 355 0 1
```


Data (in Stata)

```
. import delimited "aids.txt", delimiter(space)
  case(preserve)
. list in 1/10

     +-------------+
     |   Z   X   Y |
     |-------------|
  1. | 405   0   1 |
  2. | 412   0   0 |
  3. | 301   1   0 |
  4. | 253   1   0 |
  5. | 307   0   1 |
     |-------------|
  6. | 392   0   1 |
  7. | 361   0   0 |
  8. | 363   1   0 |
  9. | 267   1   0 |
 10. | 355   0   1 |
     +-------------+
```


Crude association (in R)

```
> chisq.test(x=aids$X, y=aids$Y)

        Pearson's Chi-squared test with Yates' continuity cor

data:  aids$X and aids$Y
X-squared = 337.47, df = 1, p-value < 2.2e-16
```

- *Interpretation?*

Crude association (in Stata)

```
. tabulate X Y, chi2

           |           Y
         X |         0          1 |     Total
-----------+----------------------+----------
         0 |       199        299 |       498
         1 |       475         27 |       502
-----------+----------------------+----------
     Total |       674        326 |     1,000

          Pearson chi2(1) = 339.9550   Pr = 0.000
```

- *Interpretation?*

The role of CD4 count

- Subjects with low CD4 count are more likely to get AZT, and more likely to get infections
- Arguable, CD4 count is an important confounder that we need to control for
- But in the data, very few subjects have the same CD4 count
  - stratification on CD4 count is not feasible
- Let's use a regression model

The logistic regression model

- Since the outcome is binary, it is natural to use the logistic regression model

$$\text{logit}\{p(Y=1|X,Z)\} = \alpha + \beta X + \gamma Z$$

- *What are the interpretations of α, β, and γ in terms of probabilities?*

Solution

$$\text{logit}\{p(Y=1|X,Z)\} = \alpha + \beta X + \gamma Z$$

$$\begin{aligned}\alpha &= \text{logit}\{p(Y=1|X=0,Z=0)\} \\ &= \log\left\{\frac{p(Y=1|X=0,Z=0)}{p(Y=0|X=0,Z=0)}\right\}\end{aligned}$$

$$\begin{aligned}\beta &= \text{logit}\{p(Y=1|X=1,Z)\} - \text{logit}\{p(Y=1|X=0,Z)\} \\ &= \log\left\{\frac{p(Y=1|X=1,Z)}{p(Y=0|X=1,Z)} \Big/ \frac{p(Y=1|X=0,Z)}{p(Y=0|X=0,Z)}\right\}\end{aligned}$$

$$\begin{aligned}\gamma &= \text{logit}\{p(Y=1|X,Z+1)\} - \text{logit}\{p(Y=1|X,Z) \\ &= \log\left\{\frac{p(Y=1|X,Z+1)}{p(Y=0|X,Z+1)} \Big/ \frac{p(Y=1|X,Z)}{p(Y=0|X,Z)}\right\}\end{aligned}$$

Causal interpretation

$$\text{logit}\{p(Y=1|X,Z)\} = \alpha + \beta X + \gamma Z$$

- If we have conditional exchangeability, given Z, then β is the conditional effect of X on Y, given Z, as a log odds ratio:

$$\beta = \log\left\{\frac{p(Y_1=1|Z)}{p(Y_1=0|Z)} \Big/ \frac{p(Y_0=1|Z)}{p(Y_0=0|Z)}\right\}$$

Underlying assumptions

$$\text{logit}\{p(Y=1|X,Z)\} = \alpha + \beta X + \gamma Z$$

- *What assumptions does this model make?*

Solution

$$\text{logit}\{p(Y=1|X,Z)\} = \alpha + \beta X + \gamma Z$$

- No effect modification by (interaction with) Z
  - The increase in log odds of being infection free, comparing AZT with no AZT, at a given CD4 count Z, is assumed to be constant ($= \beta$) across levels of Z
- Linear 'effect' of CD4 count
  - The increase in log odds for being infection free, comparing CD4 count $Z+1$ with CD4 count Z, at a given AZT level X, is assumed to be constant ($= \gamma$) across levels Z

Fundamental limitation of models

- **All models are wrong**
  - but if the model is approximately correct, then our conclusions are approximately valid
- Assumptions that we make should ideally be justified by both
  - subjects matter knowledge, and
  - data (e.g. diagnostic tests)

Fitting the model (in R)

```
> fit <- glm(formula=Y~X+Z, family=binomial,
  data=aids)
> summary(fit)

             Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.251557   0.707103   3.184  0.00145 **
X           -3.513298   0.240476 -14.610  < 2e-16 ***
Z           -0.004962   0.001882  -2.637  0.00836 **
```

- *Interpretation?*

Fitting the model (in Stata)

```
. logistic Y X Z, coef
------------------------------------------------------------------
           Y |      Coef.   Std. Err.      z    P>|z|     [95% Con
-------------+----------------------------------------------------
           X |  -3.513298   .2404814   -14.61   0.000    -3.984633
           Z |  -.0049621   .0018817    -2.64   0.008    -.0086502
       _cons |   2.251557   .7071069     3.18   0.001     .8656534
------------------------------------------------------------------
```

- *Interpretation?*

A closer look at the model (in R)

- Adding an interaction term between X and Z gives:

```
> fit <- glm(formula=Y~X+Z+X*Z, family=binomial,
   data=aids)
> summary(fit)

             Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.987432   0.766467   1.288 0.197645
X            3.309358   1.734947   1.907 0.056460 .
Z           -0.001564   0.002049  -0.763 0.445355
X:Z         -0.021647   0.005667  -3.820 0.000133 ***
```

- *Interpretation? Is the treatment beneficial or harmful?*

A closer look at the model (in Stata)

- Adding an interaction term between X and Z gives:

```
. logistic Y X Z c.X#c.Z, coef
------------------------------------------------------------------
           Y |      Coef.   Std. Err.      z    P>|z|     [95% Con
-------------+----------------------------------------------------
           X |   3.309358   1.734994     1.91   0.056    -.0911686
           Z |  -.0015638   .0020491    -0.76   0.445    -.0055799
             |
     c.X#c.Z |  -.0216467    .005667    -3.82   0.000    -.0327538
             |
       _cons |   .9874315    .766467     1.29   0.198    -.5148162
------------------------------------------------------------------
```

- *Interpretation? Is the treatment beneficial or harmful?*

What to report? (in R)

```
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.987432   0.766467   1.288 0.197645
X            3.309358   1.734947   1.907 0.056460 .
Z           -0.001564   0.002049  -0.763 0.445355
X:Z         -0.021647   0.005667  -3.820 0.000133 ***
```

- The conditional effect of X on Y, given Z, depends on Z
- Should we report the main effect together with the interaction term?
  - unintuitive for non-statisticians
  - cumbersome if many covariates and interaction terms
- Or perhaps report the effect at the mean/median of Z?
  - not very informative, unless most subjects are close to the mean/median

What to report? (in Stata)

```
------------------------------------------------------------------
           Y |      Coef.   Std. Err.      z    P>|z|     [95% Con:
-------------+----------------------------------------------------
           X |   3.309358   1.734994     1.91   0.056    -.0911686
           Z |  -.0015638   .0020491    -0.76   0.445    -.0055799
             |
     c.X#c.Z |  -.0216467    .005667    -3.82   0.000    -.0327538
             |
       _cons |   .9874315    .766467     1.29   0.198    -.5148162
------------------------------------------------------------------
```

- The conditional effect of X on Y, given Z, depends on Z
- Should we report the main effect together with the interaction term?
  - unintuitive for non-statisticians
  - cumbersome if many covariates and interaction terms
- Or perhaps report the effect at the mean/median of Z?
  - not very informative, unless most subjects are close to the mean/median

The marginal effect

$$p(Y_0 = 1) \text{ vs } p(Y_1 = 1)$$

- Arguably more intuitive than main effect + interaction term
- Can always be presented as one single number (e.g. one log odds ratio) regardless of the number of interactions
- More informative than the effect at the mean/median Z, since it applies to the whole population

The standardization formula

- If we have conditional exchangeability, given Z, then $p(Y_x = 1)$ can be computed with the standardization formula

$$p(Y_x = 1) = \sum_Z p(Y = 1|X = x, Z)p(Z)$$

- If both X and Z are binary, then we can estimate $p(Y = 1|X, Z)$ and $p(Z)$ without modeling assumptions
  - non-parametric standardization
- If X and/or Z is continuous (or categorical with many levels), non-parametric standardization is not feasible
- But we can use a regression model to estimate $p(Y_x = 1)$

Four steps for regression standardization

- **Step 1**: fit a regression model for the outcome
- **Step 2**: replace the factual exposure level with x, for each individual
- **Step 3**: estimate $p(Y = 1|X = x, Z)$ for each individual (i.e. for each observed value of Z)
- **Step 4**: average these estimates over all individuals to obtain an estimate of $p(Y_x = 1)$

Code for standardization (in R)

```
> #step 1
> fit <- glm(formula=Y~X+Z+X*Z, family=binomial,
     data=aids)
> #step 2 for x=0
> aids0 <- aids
> aids0$X <- 0
> #step 3 for x=0
> pred0 <- predict(object=fit, newdata=aids0,
   type="respons")
> #step 4 for x=0
> p0 <- mean(pred0)
> p0
[1] 0.6083575
```


Code for standardization, cont'd (in R)

```
> #step 1
> fit <- glm(formula=Y~X+Z+X*Z, family=binomial,
     data=aids)
> #step 2 for x=1
> aids1 <- aids
> aids1$X <- 1
> #step 3 for x=1
> pred1 <- predict(object=fit, newdata=aids1,
   type="respons")
> #step 4 for x=1
> p1 <- mean(pred1)
> p1
[1] 0.03748992
```


Code for standardization (in Stata)

```
. *step 1
. logistic Y X Z c.X#c.Z
. *step 2 for x=0
. replace X = 0
. *step 3 for x=0
. predict pred0
. *step 2 for x=1
. replace X = 1
. *step 3 for x=1
. predict pred1
. *step 4
. mean pred0 pred1
--------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
       pred0 |   .6083575   .0005884      .6072027    .6095122
       pred1 |   .0374899   .0014004      .0347419    .0402379
--------------------------------------------------------------
```


The marginal causal log odds ratio

$$\hat{p}(Y_0 = 1) = 0.6083575 \qquad \hat{p}(Y_1 = 1) = 0.03748992$$

- We can use the estimates of $p(Y_0 = 1)$ and $p(Y_1 = 1)$ to construct an estimate of the marginal causal log odds ratio

$$\log\left\{\frac{\hat{p}(Y_1 = 1)}{1 - \hat{p}(Y_1 = 1)} \Big/ \frac{\hat{p}(Y_0 = 1)}{1 - \hat{p}(Y_0 = 1)}\right\} = -3.68$$

- *Interpretation?*

Standard errors

- Standard errors can be obtained with some additional programming
  - sandwich formula
  - bootstrap
- Bootstrap:

$$s.e = 0.23$$

- 95% CI:

$$\text{estimate} \pm 1.96 \times s.e. = -3.68 \pm 1.96 \times 0.23 = (-3.23, -4.13)$$

Other measures of marginal effects

$$\text{logit}\{p(Y = 1|X, Z)\} = \alpha + \beta X + \gamma Z + \psi XZ$$

$$\hat{p}(Y_0 = 1) = 0.6083575 \qquad \hat{p}(Y_1 = 1) = 0.03748992$$

- Once we have estimated $p(Y_1 = 1)$ and $p(Y_0 = 1)$ separately, we can estimate any measure of effect, e.g.

$$\text{causal risk difference} = \hat{p}(Y_1 = 1) - \hat{p}(Y_0 = 1) = -0.57$$

$$\text{causal risk ratio} = \hat{p}(Y_1 = 1)/\hat{p}(Y_0 = 1) = 0.06$$

even though the estimates were derived from a logistic regression model

The package `stdReg` (in R)

```
> library(stdReg)
> fit <- glm(formula=Y~X+Z+X*Z, family=binomial,
     data=aids)
> std.fit <- stdGlm(fit=fit, data=aids, X="X")
> summary(std.fit)

Formula: Y ~ X + Z + X * Z
Family: binomial
Link function: logit
Exposure:  X

  Estimate Std. Error lower 0.95 upper 0.95
0   0.6084    0.02410     0.5611     0.6556
1   0.0375    0.00729     0.0232     0.0518
```


The package `stdReg`, cont'd (in R)

```
> summary(object=std.fit, transform="logit",
   contrast="difference", reference=0)

Formula: Y ~ X + Z + X * Z
Family: binomial
Link function: logit
Exposure:  X
Transform:  logit
Reference level:  X = 0
Contrast:  difference

  Estimate Std. Error lower 0.95 upper 0.95
0     0.00      0.000       0.00       0.00
1    -3.69      0.226      -4.13      -3.24
```

References for `stdReg`

Sjölander A. (2016). Regression standardization with the R package stdReg. *European Journal of Epidemiology* **31**(6), 563-574.
Sjölander, A. (2018). Estimation of causal effect measures with the R-package stdReg. *European Journal of Epidemiology* **33**(9), 847-858.

The `margins` command (in Stata)

```
. margins, at(X=0 X=1)
------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|
-------------+----------------------------------------------
         _at |
          1  |   .6083575   .0240991    25.24   0.000
          2  |   .0374899   .0072697     5.16   0.000
------------------------------------------------------------
```


The `margins` command, cont'd (in Stata)

```
. margins, at(X=(0 1)) post
. nlcom (log_odds_diff:
  log((_b[2._at]/(1-_b[2._at])))-
  log((_b[1._at]/(1-_b[1._at]))))
------------------------------------------------------------
              |      Coef.   Std. Err.      z    P>|z|     [9
--------------+---------------------------------------------
log_odds_diff |  -3.685885   .2254293   -16.35   0.000    -4.
------------------------------------------------------------
```


Outline

Estimation without regression models
- Standardization
- Inverse probability weighting

Estimation with regression models
- Standardization
- Inverse probability weighting

Marginal effect through exposure model: IPW

- We have seen that marginal causal effect can be estimated with IPW
- Like standardization, IPW requires modeling assumption when X and/or Z is continuous (or categorical with many levels)
- However, whereas standardization requires a model for the outcome, IPW requires a model for the exposure

Three steps for IPW

- **Step 1**: fit a regression model for the exposure
- **Step 2**: use the fitted exposure model to estimate the subject-specific weight

$$W_i = \frac{1}{p(X_i|Z_i)}$$

- **Step 3**: Use $p(Y = 1|X = x)$ in the weighted sample as an estimate of $p(Y_x = 1)$

Code for IPW (in R)

```
> #step 1
> fit <- glm(formula=X~Z, family=binomial,
  data=aids)
> #step 2
> pred <- predict(object=fit, type="respons")
> w <- 1/(aids$X*pred+(1-aids$X)*(1-pred))
> #step 3 for x=0
> p0 <- weighted.mean(aids$Y[aids$X==0], w[aids$X==0])
> p0
[1] 0.6079702
> #step 3 for x=1
> p1 <- weighted.mean(aids$Y[aids$X==1], w[aids$X==1])
> p1
[1] 0.03633255
```


Code for IPW (in Stata)

```
. *step 1
. logistic X Z
. *step 2
. predict pred
. gen w = 1/(X*pred+(1-X)*(1-pred))
. *step 3
. mean Y [pweight = w], over(X)
--------------------------------------------------------------
        Over |       Mean   Std. Err.     [95% Conf. Int
-------------+------------------------------------------------
Y            |
           0 |   .6079702   .0243257      .5602348    .6
           1 |   .0363326   .0071898      .0222237    .0
--------------------------------------------------------------
```

The marginal causal log odds ratio

$$\hat{p}(Y_0 = 1) = 0.6079702 \qquad \hat{p}(Y_1 = 1) = 0.03633255$$

- We can use the estimates of $p(Y_0 = 1)$ and $p(Y_1 = 1)$ to construct an estimate of the marginal causal log odds ratio

$$\log\left\{\frac{\hat{p}(Y_1 = 1)}{1 - \hat{p}(Y_1 = 1)} \Big/ \frac{\hat{p}(Y_0 = 1)}{1 - \hat{p}(Y_0 = 1)}\right\} = -3.72$$

- *Interpretation?*

Standard errors

- Standard errors can be obtained with some additional programming
  - sandwich formula
  - bootstrap
- Bootstrap:

$$s.e = 0.23$$

- 95% CI:

$$\text{estimate} \pm 1.96 \times s.e. = -3.72 \pm 1.96 \times 0.23 = (-3.27, -4.17)$$

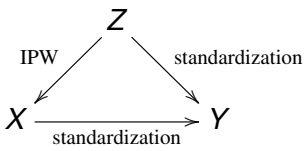
The `teffects` command (in Stata)

```
. teffects ipw (Y) (X Z), pomeans
------------------------------------------------------------
             |               Robust
           Y |      Coef.   Std. Err.      z    P>|z|     [95
-------------+----------------------------------------------
POmeans      |
           X |
          0  |   .6079702    .024303    25.02   0.000     .56
          1  |   .0363326   .0070397     5.16   0.000      .0
------------------------------------------------------------
```


Standardization vs. IPW

- Without modeling assumptions, standardization and IPW give the same results
- When using models, standardization and IPW may give different results
- What method is best?

Choice of modeling assumptions



- Standardization and IPW models different parts of the DAG
  - Standardization models how the outcome depends on the exposure and confounders
  - IPW models how the exposure depends on the confounders
- In some scenarios we may know more about one mechanism than the other, so that one model is easier to well specify
  - e.g. we may know more about the guidelines for AZT administration than we know about the biological mechanisms underlying infection

Statistical precision

- Standardization always gives more precise estimates than IPW
  - e.g. smaller standard errors and narrower confidence intervals
- The difference may be large, in particular when the exposure is continuous
  - in which case IPW requires inverse weighting with a probability density, which may give very unstable estimates
- This is an important advantage of standardization

Marginal structural models

- Both standardization and IPW can be generalized for longitudinal studies with time-varying exposures
- However, IPW is easier to use than standardization when the exposure is time-varying
  - indeed, time varying exposures was the original motivation for IPW
- When the exposure is time-varying, IPW is used to estimate causal parameters in **marginal structural models**

Doubly robust estimation

- Standardization gives an unbiased estimate if the outcome model is correct
- IPW gives an unbiased estimate if the exposure model is correct
- But if either model is incorrect, then the obtained estimate is generally biased
- It is possible to combine both models into a **doubly robust** estimator
  - unbiased if either model is correct, not necessarily both
  - two chances of valid inference instead of only one
  - beyond the scope of this course

Summary

- Standard methods for confounding control (e.g. stratification and outcome regression models) give conditional (subpopulation) causal effects
  - provided that we have conditional exchangeability, given the measured confounders
- Often, marginal (population) causal effect may be more relevant target parameters
- Marginal causal effects can be estimated with standardization and IPW
  - either non-parametrically, or with regression models