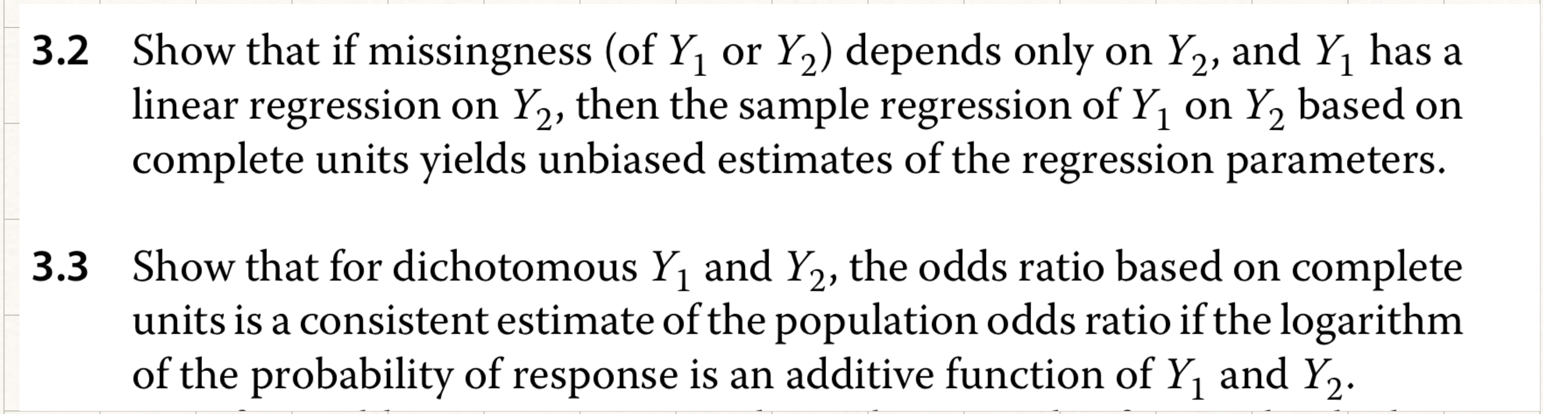

3.2 Show that if missingness (of Y_1 or Y_2) depends only on Y_2, and Y_1 has a linear regression on Y_2, then the sample regression of Y_1 on Y_2 based on complete units yields unbiased estimates of the regression parameters.

3.3 Show that for dichotomous Y_1 and Y_2, the odds ratio based on complete units is a consistent estimate of the population odds ratio if the logarithm of the probability of response is an additive function of Y_1 and Y_2.

3.2: $Y_1 = \beta_0 + \beta_1 Y_2 + \varepsilon$

Denote $\gamma = (\beta_0, \beta_1)$
$Z_i = (1, Y_{2i})$

Now Denote that $R_i = \begin{cases} 1 & Y_i \text{ is not missing} \\ 0 & Y_i \text{ is missing} \end{cases}$

Then we need ~~to~~ solve the OLS problem

$$\min \sum_{i=1}^{n} R_i (Y_{1i} - \beta_0 - \beta_1 Y_{2i})^2 = \min \sum_{i=1}^{n} R_i (Y_{1i} - \gamma^T Z_i)$$

By Lecture notes: $\hat{\gamma} = E(RZ^T Z)^{-1} E(RZ^T Y_1)$

Now At MAR mechanism, that $R \perp\!\!\!\perp Y_1 \mid Y_2$

$$\hat{\gamma}' = E(RZ^T Z)^{-1} E(RZ^T(\gamma' Z + \varepsilon))$$

$\downarrow$

as $n \to +\infty$

$2 \times 1 \times 1 \times 2 \times 2 \times 1$

Notice that $\mathbb{E}(RZ^T(\gamma' Z+\varepsilon))$

$= \mathbb{E}(RZ^TZ)\gamma + \mathbb{E}(RZ^T\varepsilon)$

Notice $R \perp\!\!\!\perp \varepsilon \mid Z \Rightarrow \mathbb{E}(RZ^T(\gamma' Z+\varepsilon)) = \mathbb{E}(RZ^TZ)\gamma'$

$\Rightarrow$ as $n \to +\infty$

$$\hat{\gamma}' = \mathbb{E}(RZ^TZ)^{-1} \cdot \mathbb{E}(RZ^TZ)\gamma = \gamma$$

$\therefore \hat{\gamma}'$ is unbiased

$\Downarrow$

which also implies

3·3

3.3 Show that for dichotomous Y_1 and Y_2, the odds ratio based on complete units is a consistent estimate of the population odds ratio if the logarithm of the probability of response is an additive function of Y_1 and Y_2.

		1	0
Y_1	1	(1,1)	(1,0)
	0	(0,1)	(0,0)

Y_2

suppose the probability of response can be written as $P(Y_1, Y_2) = \alpha + \beta_1 Y_1 + \beta_2 Y_2$

$$\therefore P(Y_1=1, Y_2=1) = \alpha+\beta_1+\beta_2$$

$$P(Y_1=0, Y_2=1) = \alpha+\beta_2$$

$$P(Y_1=1, Y_2=0) = \alpha+\beta_1$$

$$P(Y_1=0, Y_2=0) = \alpha$$

$\therefore$ The true odd ratio should be

$$\frac{(\alpha+\beta_1+\beta_2)\alpha}{(\alpha+\beta_2)(\alpha+\beta_1)} = \frac{\alpha^2+\beta_1\alpha+\beta_2\alpha}{\alpha^2+\alpha\beta_1+\alpha\beta_2+\beta_1\beta_2}$$

Suppose we have

		Y_2	
		1	0
Y_1	1		
	0		

Suppose we have totally n observations, there are i numbers of missing Y_1 and j numbers of missing Y_2. , and k number of both Y_1 and Y_2 missing.

Then

		Y_2	
		1	0
Y_1	1		
	0		

fraction of $Y_1=1, Y_2=1$ missing is $\frac{i+j}{n}$

That in fact, we will loss $\frac{i+j}{n}$ of data suppose that there are A numbers of $Y_1 = 1$, then there are $n - i - A$ number of $Y_1 = 0$

suppose there are m numbers of $Y_2 = 1$,

Then there are $n - j - m$ number of $Y_2 = 0$

comment: I really don't know how to show this problem, that I don't know what is the loss function, and how to estimate $P(Y_1=1, Y_2=1)$, $P(Y_1=0, Y_2=1)$, ...

I don't know how to formulate the relationship between loss function and the odd ratio.