Text Exercise 4

Part (a)

The answer to the third question (d_i) might depend on ommited factors. One possibility would be following another diet. Following another diet would affect both variable d_i and the change in weight ($y_{i1} - y_{i0}$). This can result in an underestimation of significance of d_i. If the participant follow another diet, d_i will be zero and ($y_{i1} - y_{i0}$) might be bigger than zero.

Part (b)

The two conditions can be rephrased as follows:

1. $\frac{1}{n}Z'\varepsilon \to 0$: The variable z representing the door-to-door advertising is not correlated to the residual of the OLS model in Part (a).
2. $\frac{1}{n}Z'X \to Q \neq 0$: The variable z representing the door-to-door advertising is sufficiently correlated to the residual of the OLS model in Part (a).

Part (c)

1. $\frac{1}{n}Z'\varepsilon$: possible to test this statistically using Sargan test.
2. $\frac{1}{n}Z'X$: possible to test statistically. Use t-test for coefficient significance for OLS with formula $X = \gamma Z + \eta$ where η is the error term.

Part (d)

$$\beta = (Z'X)^{-1}(Z'Y)$$

$$Z = \begin{pmatrix} 1 & z_1 \\ 1 & z_2 \\ \vdots & \\ 1 & z_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & d_1 \\ 1 & d_2 \\ \vdots & \\ 1 & d_n \end{pmatrix} \quad Y = \begin{pmatrix} (y_{11} - y_{10}) \\ (y_{21} - y_{20}) \\ \vdots \\ (y_{n1} - y_{n0}) \end{pmatrix}$$

$$Z'X = \begin{pmatrix} 1 & 1 & ... & 1 \\ z_1 & z_2 & ... & z_n \end{pmatrix} \begin{pmatrix} 1 & d_1 \\ 1 & d_2 \\ \vdots & \\ 1 & d_n \end{pmatrix} = \begin{pmatrix} n \\ \sum_{i=1}^n z_i d_i \end{pmatrix}$$

$$(Z'X)^{-1} = \begin{pmatrix} \frac{1}{n} & \frac{1}{\sum_{i=1}^n z_i d_i} \end{pmatrix}$$

$$Z'Y = \begin{pmatrix} 1 & 1 & ... & 1 \\ z_1 & z_2 & ... & z_n \end{pmatrix} \begin{pmatrix} (y_{11} - y_{10}) \\ (y_{21} - y_{20}) \\ \vdots \\ (y_{n1} - y_{n0}) \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n (y_{i1} - y_{i0}) \\ \sum_{i=1}^n z_i (y_{i1} - y_{i0}) \end{pmatrix}$$

$$\begin{aligned}
\therefore \beta &= \begin{pmatrix} \frac{1}{n} & \frac{1}{\sum_{i=1}^n z_i d_i} \end{pmatrix} \begin{pmatrix} \sum_{i=1}^n (y_{i1} - y_{i0}) \\ \sum_{i=1}^n z_i (y_{i1} - y_{i0}) \end{pmatrix} \\
&= \frac{\sum_{i=1}^n (y_{i1} - y_{i0})}{n} + \frac{\sum_{i=1}^n z_i (y_{i1} - y_{i0})}{\sum_{i=1}^n z_i d_i} \\
&= \Delta + \frac{\sum\limits_{k \in \{i \mid z_i = 1\}} z_i (y_{i1} - y_{i0}) + \overbrace{\sum\limits_{k \in \{i \mid z_i = 0\}} z_i (y_{i1} - y_{i0})}^{=0}}{\sum\limits_{k \in \{i \mid z_i = 1\}} z_i d_i + \underbrace{\sum\limits_{k \in \{i \mid z_i = 0\}} z_i d_i}_{=0}} \\
&= \Delta + \frac{\sum_{k \in \{i \mid z_i = 1\}} z_i (y_{i1} - y_{i0})}{\sum_{k \in \{i \mid z_i = 1\}} z_i d_i} \\
&= \Delta + \frac{\Delta^1}{\bar{d}^1}
\end{aligned}$$