a). ① People can choose whether to follow the diet in the year past as they wish.

② Other variables, except for those already included in the vector x_i and y_{i0}, exist while affecting both d_i and $(y_{i1} - y_{i0})$. For example:
External education on healthy life style (keeping fit and a healthy diet).

b) ① The instrument Z should not be significantly correlated with the dependent variable $(y_{i1} - y_{i0})$. In other words, whether the regions have door-to-door advertising or not makes no difference in terms of the weight conditions.

② The instrument Z should significantly correlate with the explanatory variable d_i. In other words, people in regions with door-to-door advertising are more likely to follow the diet.

c). ① For $\frac{1}{n} Z'\varepsilon \to 0$ when $n \to \infty$:

No. Sargan test requires that the number of instruments m ($m=1$ in this case) to be larger than the number of explanatory variables k ($k=3$). When $m > k$ is violated, we cannot tell whether the instrument is valid.

② For $\frac{1}{n} Z'X \to Q$ ($Q \neq 0$) when $n \to \infty$:

Yes. Regress d_i on z_i, y_{i0} and x_i', and test if z_i significantly correlates with d_i.

d). $Z = \begin{pmatrix} 1 & z_1 \\ 1 & z_2 \\ \vdots & \vdots \\ 1 & z_n \end{pmatrix}$, $X = \begin{pmatrix} 1 & d_1 \\ 1 & d_2 \\ \vdots & \vdots \\ 1 & d_n \end{pmatrix}$, then $(Z'X)^{-1}Z'y = \begin{pmatrix} n & \sum d_i \\ \sum z_i & \sum d_i z_i \end{pmatrix}\begin{pmatrix} \sum y_i \\ \sum z_i y_i \end{pmatrix}$

$$= \frac{1}{n\sum d_i z_i - \sum z_i \sum d_i}\begin{pmatrix} \sum d_i z_i & -\sum d_i \\ -\sum z_i & n \end{pmatrix}\begin{pmatrix} \sum y_i \\ \sum z_i y_i \end{pmatrix}$$

$$b = \underset{2SLS}{\hat{\beta}} = \frac{n\sum z_i y_i - \sum y_i \sum z_i}{n\sum d_i z_i - \sum z_i \sum d_i} = \frac{\sum z_i y_i - \frac{\sum y_i \sum z_i}{n}}{\sum d_i z_i - \frac{\sum z_i \sum d_i}{n}} = \frac{\frac{1}{\sum z_i}\sum z_i y_i - \frac{\sum y_i}{n}}{\frac{1}{\sum z_i}\sum d_i z_i - \frac{\sum d_i}{n}} = \frac{\Delta' - \Delta}{d' - d}$$

Kunyu HE