Answers to HW2

Problem 3.2.
From \( pr[\mathbf{m} = (0,0)|Y_1,Y_2] = pr[\mathbf{m} = (0,0)|Y_2] \), we have \( pr[Y_1|Y_2,\mathbf{m} = (0,0)] = pr[Y_1|Y_2] \). Then \( E[Y_1|Y_2,\mathbf{m} = (0,0)] = E[Y_1|Y_2] \) and \( var[Y_1|Y_2,\mathbf{m} = (0,0)] = var[Y_1|Y_2] \). Hence the sample regression of \( Y_1 \) on \( Y_2 \) based on complete cases yields unbiased estimates of the regression parameters of \( [Y_1|Y_2,\mathbf{m} = (0,0)] \) and \( [Y_1|Y_2] \).

Problem 3.3.
Assume that \( [n_{ij}|N_{ij}] \sim Bin(N_{ij},p_{ij}) \), where \( p_{ij} = exp\{g_1(i) + g_2(j)\} \). Then as \( \{N_{ij},i,j = 1,2\} \to \infty \),

\[
\frac{n_{11}n_{22}}{n_{12}n_{21}} = \frac{\frac{n_{11}}{N_{11}}\frac{n_{22}}{N_{22}}N_{11}N_{22}}{\frac{n_{12}}{N_{12}}\frac{n_{21}}{N_{21}}N_{12}N_{21}} \to \frac{N_{11}N_{22}}{N_{12}N_{21}}.
\]

Problem 3.4.

\[
\bar{y} = \frac{1}{r} \sum_{i=1}^{r} y_i = \frac{22 \times 220 + 27 \times 225 + 16 \times 250 + 5 \times 270}{22 + 27 + 16 + 5} = 232.36.
\]

s.e.(\( \bar{y} \)) = \( \sqrt{\frac{1}{r-1} \sum_{i=1}^{r} (y_i - \bar{y})^2} = \sqrt{\frac{1}{r(r-1)} \left( \sum_{i=1}^{r} y_i^2 - r \bar{y}^2 \right)} \)

\[
= \sqrt{\frac{1}{r(r-1)} \left( \sum_{j=1}^{4} \sum_{i=1}^{r_j} y_{ji}^2 - r \bar{y}^2 \right)} = \sqrt{\frac{1}{r(r-1)} \left( \sum_{j=1}^{4} \left( \sum_{i=1}^{r_j} (y_{ji} - \bar{y}_{jR})^2 + r_j \bar{y}_{jR}^2 \right) - r \bar{y}^2 \right)}
\]

\[
= \sqrt{\frac{1}{r(r-1)} \sum_{j=1}^{4} \left( (r_j - 1)SD_j^2 + r_j \bar{y}_{jR}^2 \right) - r \bar{y}^2]}
\]

\[
\approx 4.62.
\]

Therefore the 95% CI for \( \bar{y} \) is (223.30,241.42). This interval can not be applied to all individuals in the county since the responders do not seem like a random sample of the entire population according to the discrepancy among the stratum-specific responding probabilities.

Problem 3.5. The weighting class estimate is

\[
\bar{y}_{wc} = \frac{1}{n} \sum_{j=1}^{J} n_j \bar{y}_{jR} = \frac{25 \times 220 + 35 \times 225 + 28 \times 250 + 12 \times 270}{100} = 236.15.
\]
While using (3.7) to calculate its m.s.e., we can assume $\frac{n}{N} \approx 0$. After all, we have $m.s.e.(\bar{y}_{wc}) \approx 25.44$. Then the 95% CI is about (226.26, 246.04). Compared to the CI for $\bar{y}$, the CI for $\bar{y}_{wc}$ is slightly shifted to the right and a little bit wider. The *quasirandomization* is assumed on the nonresponse mechanism for $\bar{y}_{wc}$.

**Problem 3.7** Use formula (3.4) to calculate the HT estimator. In order to get a weighting class estimator, you may stratify the subjects, for example according to the values of $x$: \{(1,2,3),(4,5,6),(7,8,9,10)\}. Then use formula (3.5) to calculate the weighting class estimator.