Midterm: STAT 570

Philip Pham

November 12, 2018

1. Consider an situation in which we are interested in the risk of death in the first 5 years of life (the under-5 mortality mortality risk, or U5MR) in each of $2n$ areas in two consecutive time periods. Consider a hypothetical situation in which a malaria prevention intervention is randomized across the areas, immediately after the first time periods. Areas indexed by $i = 1, \ldots, n$ are control areas, while areas $i = n + 1, \ldots, 2n$ receiving the intervention.

   In each area and each time period alive/dead status of M_{it} children are recorded, call the number dead D_{it} for $i = 1, \ldots, 2n$, $t = 0, 1$. Let

   $$Y_{it} = \log\left(\frac{D_{it}/M_{it}}{1 - D_{it}/M_{it}}\right), \tag{1}$$

   denote the logit of the U5MR in area i in period t, $i = 1, \ldots, n$, $t = 0, 1$.

   Suppose the true model is given by

   $$Y_{it} = \beta_0 + \alpha_i + \beta_1 x_{it} + \epsilon_{it}, \tag{2}$$

   where $\alpha_i \sim \mathcal{N}\left(0, \sigma_\alpha^2\right)$ are area-specific random effects and $\epsilon_{it} \sim \mathcal{N}\left(0, \sigma_\epsilon^2\right)$, represents measurement error, with α_i and ϵ_{it} independent, $i = 1, \ldots, 2n$, $t = 0, 1$. The covariate x_{it} is an indicator for the intervention so that $x_{i0} = 0$ for $i = 1, \ldots, 2n$, $x_{i1} = 0$ for $i = 1, \ldots, n$, and $x_{i1} = 1$ for $i = n + 1, \ldots, 2n$.

   We will consider three models for the child mortality data:

   Follow-up model: $Y_{i1} = \beta_0^\dagger + \beta_1^\dagger x_{i1} + \epsilon_{i1}^\dagger$, for $i = 1, \ldots, 2n$.

   Change model: $Z_i = Y_{i1} - Y_{i0} = \beta_0^\star + \beta_1^\star x_{i1} + \epsilon_i^\star$, for $i = 1, \ldots, 2n$.

   Analysis for Covariance (ANCOVA) model: $Y_{i1} = \beta_0^\ddagger + \gamma Y_{i0} + \beta_1^\ddagger x_{i1} + \epsilon_i^\ddagger$, for $i = 1, \ldots, 2n$.

   (a) Carefully interpret $\beta_1^\dagger$, $\beta_1^\star$ and $\beta_1^\ddagger$ in these models, and hence what each of $\mathbb{E}\left[\hat{\beta}_1^\dagger\right]$, $\mathbb{E}\left[\hat{\beta}_1^\star\right]$, and $\mathbb{E}\left[\hat{\beta}_1^\ddagger\right]$ are unbiased estimators of.

   Solution: Let's examine each case.

   $\beta_1^\dagger$: Let $Y_{:,1} = \begin{pmatrix} Y_{1,1} & \cdots & Y_{2n,1} \end{pmatrix}^\intercal$. Let $\beta = \begin{pmatrix} \beta_0 & \beta_1 \end{pmatrix}^\intercal$. Let X be the $2n \times 2$ matrix with 1s in the first column and $x_{1,1}, \ldots, x_{2n,1}$ in the second column. We can write $Y_{:,1} = X\beta + \alpha_i + \epsilon_{:,1}$.

We have that

$$
\begin{aligned}
\hat{\beta}^{\dagger} &= \left(X^{\intercal}X\right)^{-1} X^{\intercal}Y_{:,1} = \left(X^{\intercal}X\right)^{-1} X^{\intercal}\left(X\beta + \alpha + \epsilon_{:,1}\right) \\
&= \beta + \left(X^{\intercal}X\right)^{-1} X^{\intercal}\left(\alpha + \epsilon_{:,1}\right) \\
&\sim \mathcal{N}\left(\beta, \left(\sigma_{\alpha}^{2} + \sigma_{\epsilon}^{2}\right)\left(X^{\intercal}X\right)^{-1}\right),
\end{aligned} \tag{3}
$$

so we'll obtain unbiased estimates of β with higher variance than if we had the correct model.

So, $\beta_1^{\dagger}$ is the expected change in the logit of the U5MR after applying the treatment.

$\beta_1^{\star}$: We have that $Z_i = Y_{i1} - Y_{i0} = \beta_1\left(x_{i1} - x_{i0}\right) + \epsilon_{i1} - \epsilon_{i0} = \beta_1 x_{i1} + \left(\epsilon_{i1} - \epsilon_{i0}\right)$. Solving for $\hat{\beta}^{\star}$, we find

$$
\begin{aligned}
\hat{\beta}^{\star} &= \left(X^{\intercal}X\right)^{-1} X^{\intercal}Z_i = \left(X^{\intercal}X\right)^{-1} X^{\intercal}\left(X\begin{pmatrix}0\\ \beta_1\end{pmatrix} + \left(\epsilon_{:,1} - \epsilon_{:,0}\right)\right) \\
&= \begin{pmatrix}0\\ \beta_1\end{pmatrix} + \left(X^{\intercal}X\right)^{-1} X^{\intercal}\left(\epsilon_{:,1} - \epsilon_{:,0}\right) \\
&\sim \mathcal{N}\left(\begin{pmatrix}0\\ \beta_1\end{pmatrix}, 2\sigma_{\epsilon}^{2}\left(X^{\intercal}X\right)^{-1}\right),
\end{aligned} \tag{4}
$$

so $\hat{\beta}_1^{\star}$ is an unbiased estimate of β_1.

Thus, $\beta_1^{\star}$ is again the expected change in the logit of the U5MR after applying the treatment.

$\beta_1^{\ddagger}$: Consider the different ways of writing Y_{i1},

$$
\begin{aligned}
Y_{i1} &= \beta_0 + \alpha_i + \beta_1 x_{i1} + \epsilon_{i1} \\
&= \left(\beta_0 + \alpha_i + \beta_1 x_{i0} + \epsilon_{i0}\right) + \beta_1 x_{i1} + \epsilon_{i1} - \epsilon_{i0} \\
&= Y_{i0} + \beta_1 x_{i1} + \left(\epsilon_{i1} - \epsilon_{i0}\right) \\
&= \beta_0^{\ddagger} + \gamma Y_{i0} + \beta_1^{\ddagger} x_{i1} + \epsilon_i^{\ddagger}.
\end{aligned}
$$

Define $X^{\ddagger}$ to be the $2n \times 3$ matrix with the first two columns being X and third column being $Y_{:,0}$.

Then, we have that

$$
\begin{aligned}
\begin{pmatrix}\hat{\beta}_0^{\ddagger}\\ \hat{\beta}_1^{\ddagger}\\ \hat{\gamma}\end{pmatrix} &= \left(\left(X^{\ddagger}\right)^{\intercal} X^{\ddagger}\right)^{-1}\left(X^{\ddagger}\right)^{\intercal} Y_{:,1} \\
&= \left(\left(X^{\ddagger}\right)^{\intercal} X^{\ddagger}\right)^{-1}\left(X^{\ddagger}\right)^{\intercal}\left(X^{\ddagger}\begin{pmatrix}0\\ \beta_1\\ 1\end{pmatrix} + \epsilon_{:,1} - \epsilon_{:,0}\right) \\
&= \begin{pmatrix}0\\ \beta_1\\ 1\end{pmatrix} + \left(\left(X^{\ddagger}\right)^{\intercal} X^{\ddagger}\right)^{-1}\left(X^{\ddagger}\right)^{\intercal}\left(\epsilon_{:,1} - \epsilon_{:,0}\right) \\
&\sim \mathcal{N}\left(\begin{pmatrix}0\\ \beta_1\\ 1\end{pmatrix}, 2\sigma_{\epsilon}^{2}\left(\left(X^{\ddagger}\right)^{\intercal} X^{\ddagger}\right)^{-1}\right).
\end{aligned} \tag{5}
$$

Again, $\hat{\beta}_1^{\ddagger}$ is an unbiased estimate of β_1.

All in all, we have that the expected value of the estimates

$$\mathbb{E}\left[\hat{\beta}_1^{\dagger}\right] = \mathbb{E}\left[\hat{\beta}_1^{\star}\right] = \mathbb{E}\left[\hat{\beta}_1^{\ddagger}\right] = \beta_1, \tag{6}$$

so $\beta_1^{\dagger}$, $\beta_1^{\star}$, $\beta_1^{\ddagger}$ can all be interpreted as the expected change in U5MR after applying the treatment.

(b) Evaluate $\operatorname{var}\left(\hat{\beta}_1^{\dagger}\right)$, $\operatorname{var}\left(\hat{\beta}_1^{\star}\right)$, and $\operatorname{var}\left(\hat{\beta}_1^{\ddagger}\right)$. Comment on the efficiency of the estimators arising from each of the three models.

Solution: While Equation 6 tells us that the expectation of our estimators is the same, the variances are different.

$\hat{\beta}_1^{\dagger}$: We can compute the variance from Equation 3. First, we have that

$$\begin{aligned} X^{\intercal}X &= \begin{pmatrix} 2n & \sum_{i=1}^{2n} x_{i1} \\ \sum_{i=1}^{2n} x_{i1} & \sum_{i=1}^{2n} x_{i1}^2 \end{pmatrix} = \begin{pmatrix} 2n & n \\ n & n \end{pmatrix} \\ \implies (X^{\intercal}X)^{-1} &= \frac{1}{n^2}\begin{pmatrix} n & -n \\ -n & 2n \end{pmatrix} = \frac{1}{n}\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}. \end{aligned} \tag{7}$$

Thus, we find that

$$\operatorname{var}\left(\hat{\beta}_1^{\dagger}\right) = \frac{2}{n}\left(\sigma_\alpha^2 + \sigma_\epsilon^2\right). \tag{8}$$

$\hat{\beta}_1^{\star}$: Using Equations 4 and 7, we compute that

$$\operatorname{var}\left(\hat{\beta}_1^{\star}\right) = \frac{4}{n}\sigma_\epsilon^2. \tag{9}$$

$\hat{\beta}_1^{\ddagger}$: We use Equation 5 to compute the variance. First, we note that

$$\begin{aligned} \operatorname{var}\left(\hat{\beta}_1^{\ddagger}\right) &= 2\sigma_\epsilon^2\left(\left(X^{\ddagger}\right)^{\intercal}X^{\ddagger}\right)^{-1}_{2,2} \\ &= 2\sigma_\epsilon^2\left(\frac{2n\sum_{i=1}^{2n} Y_{i0}^2 - \left(\sum_{i=1}^{2n} Y_{i0}\right)^2}{\det\left(\left(X^{\ddagger}\right)^{\intercal}X^{\ddagger}\right)}\right), \end{aligned} \tag{10}$$

where

$$\begin{aligned} \det\left(\left(X^{\ddagger}\right)^{\intercal}X^{\ddagger}\right) = \; & n\left(2n\sum_{i=1}^{2n} Y_{i0}^2 - \left(\sum_{i=1}^{2n} Y_{i0}\right)^2\right) \\ & - n\det\left(\sum_{i=1}^{2n}\begin{pmatrix} x_{i1} & x_{i1}Y_{i0} \\ Y_{i0} & Y_{i0}^2 \end{pmatrix}\right) \\ & - \left(\sum_{i=1}^{2n} x_{i1}Y_{i0}\right)\det\left(\sum_{i=1}^{2n}\begin{pmatrix} 1 & Y_{i0} \\ x_{i1} & x_{i1}Y_{i0} \end{pmatrix}\right). \end{aligned} \tag{11}$$

Note that $\det\left(\left(X^{\ddagger}\right)^{\intercal}X^{\ddagger}\right) \gneq 0$ since $\left(X^{\ddagger}\right)^{\intercal}X^{\ddagger}$ is positive definite. The first term is positive since it is $4n^3$ times the MLE estimate for the variance of $Y_{:,0}$.

For the second term, note that

$$\det\left(\sum_{i=1}^{2n}\begin{pmatrix} x_{i1} & x_{i1}Y_{i0} \\ Y_{i0} & Y_{i0}^2\end{pmatrix}\right) = n\sum_{i=1}^{2n} Y_{i0}^2 - \left(\sum_{i=n+1}^{2n} Y_{i0}\right)\left(\sum_{i=1}^{2n} Y_{i0}\right) \tag{12}$$

$$= \left(n\sum_{i=1}^{n} Y_{i0}^2 - \left(\sum_{i=n+1}^{2n} Y_{i0}\right)\left(\sum_{i=1}^{n} Y_{i0}\right)\right) + \left(n\sum_{i=n+1}^{2n} Y_{i0}^2 - \left(\sum_{i=n+1}^{2n} Y_{i0}\right)^2\right),$$

so the second term is $2n^3$ times an estimator for the variance of $Y_{:,0}$.
For the third term, note that

$$\det\left(\sum_{i=1}^{2n}\begin{pmatrix} 1 & Y_{i0} \\ x_{i1} & x_{i1}Y_{i0}\end{pmatrix}\right) = 2n\sum_{i=1}^{2n} x_{i1}Y_{i0} - \left(\sum_{i=1}^{2n} x_{i1}\right)\left(\sum_{i=1}^{2n} Y_{i0}\right),$$

$$= 2n\sum_{i=n+1}^{2n} Y_{i0} - n\sum_{i=1}^{2n} Y_{i0}. \tag{13}$$

so the third term is $4n^2\sum_{i=1}^{2n} x_{i1}Y_{i0}$ times the MLE estimate for the covariance of $x_{:,1}$ and $Y_{:,0}$, which should be 0 if the treatment is randomized. Therefore, the numerator of Equation 10 is

$$\lim_{n\to\infty} \frac{2n\sum_{i=1}^{2n} Y_{i0}^2 - \left(\sum_{i=1}^{2n} Y_{i0}\right)^2}{(2n)^2} = \operatorname{var}\left(Y_{:,0}\right), \tag{14}$$

and for the denominator, we use Equations 11, 12, 13 to obtain

$$\lim_{n\to\infty} \frac{\det\left(\left(X^\ddagger\right)^\intercal X^\ddagger\right)}{(2n)^2} = \frac{n}{2}\operatorname{var}\left(Y_{:,0}\right), \tag{15}$$

so $\lim_{n\to\infty}\operatorname{var}\left(\hat{\beta}_1^\ddagger\right) = \frac{4}{n}\sigma_\epsilon^2$, which is the same as Equation 9.

Thus, we have that the follow-up model estimates β_1 most efficiently if $\sigma_\alpha^2 < \sigma_\epsilon^2$. In most cases, we'd expect that the measurement error is smaller than the random effect, that is, $\sigma_\epsilon^2 < \sigma_\alpha^2$, so the change model would estimate β_1 most efficiently in that case. Asymptotically, the ANCOVA model is just as efficient as the change model.

In practice, since we don't have an infinite number of samples of $\operatorname{var}\left(\hat{\beta}_1^\ddagger\right) \gneqq \operatorname{var}\left(\hat{\beta}_1^\star\right)$. To see this, we show that

$$\frac{1}{n}\det\left(\left(X^\ddagger\right)^\intercal X^\ddagger\right) \lneqq \frac{1}{2}\left(2n\sum_{i=1}^{2n} Y_{i0}^2 - \left(\sum_{i=1}^{2n} Y_{i0}\right)^2\right), \tag{16}$$

which would imply that $\operatorname{var}\left(\hat{\beta}^\ddagger\right) \gneqq \frac{4}{n}\sigma_\epsilon^2$ in Equation 10.

From Equations 11, 12, 13, we have that

$$\begin{aligned}
\frac{1}{n}\det\left(\left(X^{\ddagger}\right)^{\intercal}X^{\ddagger}\right) &= n\sum_{i=1}^{2n}Y_{i0}^2-\left(\sum_{i=1}^{n}Y_{i0}\right)^2-\left(\sum_{i=1}^{n}Y_{i0}\right)\left(\sum_{i=n+1}^{2n}Y_{i0}\right)\\
&\quad-\left(\sum_{i=n+1}^{2n}Y_{i0}\right)\left(2\sum_{i=n+1}^{2n}Y_{i0}-\sum_{i=1}^{2n}Y_{i0}\right)\\
&= n\sum_{i=1}^{2n}Y_{i0}^2-\left(\sum_{i=1}^{n}Y_{i0}\right)^2-\left(\sum_{i=n+1}^{2n}Y_{i0}\right)^2. \qquad (17)
\end{aligned}$$

Using this result and substituting, we'll have that Equation 16 is true if and only if

$$n\sum_{i=1}^{2n}Y_{i0}^2-\left(\sum_{i=1}^{n}Y_{i0}\right)^2-\left(\sum_{i=n+1}^{2n}Y_{i0}\right)^2 \lneq n\sum_{i=1}^{2n}Y_{i0}^2-\frac{1}{2}\left(\sum_{i=1}^{n}Y_{i0}+\sum_{i=n+1}^{2n}Y_{i0}\right)^2.$$

With some algebra, this becomes

$$\begin{aligned}
0 &\lneq \left(\sum_{i=1}^{n}Y_{i0}\right)^2+\left(\sum_{i=n+1}^{2n}Y_{i0}\right)^2-2\left(\sum_{i=1}^{n}Y_{i0}\right)\left(\sum_{i=n+1}^{2n}Y_{i0}\right)\\
&= \left(\sum_{i=1}^{n}Y_{i0}-\sum_{i=n+1}^{2n}Y_{i0}\right)^2,
\end{aligned}$$

which is almost surely true, so we have proved Equation 16, which shows that $\operatorname{var}\left(\hat{\beta}_1^{\ddagger}\right) \gneq \frac{4}{n}\sigma_\epsilon^2 = \operatorname{var}\left(\hat{\beta}_1^{\star}\right)$, so the ANCOVA model is less efficient than the change model in general.

(c) Obtain an expression for $\hat{\gamma}$, in as simple a form as you can find.

Solution: From Equation 5, we have that

$$\begin{aligned}
\hat{\gamma} &= \left(\left(\left(X^{\ddagger}\right)^{\intercal}X^{\ddagger}\right)^{-1}\left(X^{\ddagger}\right)^{\intercal}Y_{:,1}\right)_3 \qquad (18)\\
&= \sum_{k=1}^{3}\left(\left(X^{\ddagger}\right)^{\intercal}X^{\ddagger}\right)^{-1}_{3k}\left(\left(X^{\ddagger}\right)^{\intercal}Y_{:,1}\right)_k\\
&= -n\frac{\sum_{i=1}^{n}Y_{i0}}{\det\left(\left(X^{\ddagger}\right)^{\intercal}X^{\ddagger}\right)}\sum_{i=1}^{2n}Y_{i1}+n\frac{\sum_{i=1}^{n}Y_{i0}-\sum_{i=n+1}^{2n}Y_{i0}}{\det\left(\left(X^{\ddagger}\right)^{\intercal}X^{\ddagger}\right)}\sum_{i=n+1}^{2n}Y_{i1}\\
&\quad+\frac{n^2}{\det\left(\left(X^{\ddagger}\right)^{\intercal}X^{\ddagger}\right)}\sum_{i=1}^{2n}Y_{i0}Y_{i1}\\
&= \frac{n}{\det\left(\left(X^{\ddagger}\right)^{\intercal}X^{\ddagger}\right)}\left(n\sum_{i=1}^{2n}Y_{i0}Y_{i1}-\sum_{i=1}^{n}Y_{i0}\sum_{i=1}^{n}Y_{i1}-\sum_{i=n+1}^{2n}Y_{i0}\sum_{i=n+1}^{2n}Y_{i1}\right),
\end{aligned}$$

where $n\big/\det\left(\left(X^{\ddagger}\right)^{\intercal}X^{\ddagger}\right)$ can be obtained from Equation 17.

(d) On the basis of the previous question, or otherwise, give intuitive explanations for the efficiency results in Part 1b.

Solution: Denote the MLE estimates of the covariance between Y_{i0} and Y_{i1} without and with the intervention by

$$\widehat{\text{cov}}\left(Y_{i0}, Y_{i1} \mid x_{i1} = 0\right) = \frac{1}{n}\sum_{i=1}^{n} Y_{i0}Y_{i1} - \left(\frac{1}{n}\sum_{i=1}^{n} Y_{i0}\right)\left(\frac{1}{n}\sum_{i=1}^{n} Y_{i1}\right) \tag{19}$$

$$\widehat{\text{cov}}\left(Y_{i0}, Y_{i1} \mid x_{i1} = 1\right) = \frac{1}{n}\sum_{i=n+1}^{2n} Y_{i0}Y_{i1} - \left(\frac{1}{n}\sum_{i=n+1}^{2n} Y_{i0}\right)\left(\frac{1}{n}\sum_{i=n+1}^{2n} Y_{i1}\right),$$

respectively.