# Midterm: STAT 570

Philip Pham

November 14, 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 $\alpha_i$ and $\epsilon_{it}$ 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 $\beta$ 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 $\beta_1$.
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 $Y_{i1}$,

$$\begin{aligned}Y_{i1} &= \beta_0 + \alpha_i + \beta_1 x_{i1} + \epsilon_{i1} \\ &= \beta_0^{\ddagger} + \gamma Y_{i0} + \beta_1^{\ddagger}x_{i1} + \epsilon_i^{\ddagger} \\ &= \beta_0^{\ddagger} + \beta_1^{\ddagger}x_{i1} + \gamma\left(\beta_0 + \alpha_i + \epsilon_{i0}\right) + \epsilon_i^{\ddagger}\end{aligned} \tag{5}$$

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

$$\begin{pmatrix}\hat{\beta}_0^{\ddagger}\\ \hat{\beta}_1^{\ddagger}\\ \hat{\gamma}\end{pmatrix} = \left(\left(X^{\ddagger}\right)^{\intercal} X^{\ddagger}\right)^{-1}\left(X^{\ddagger}\right)^{\intercal} Y_{:,1}. \tag{6}$$

From Equation 5, note that

$$\begin{aligned}Y_{i1} - \gamma Y_{i0} &= \left(1-\gamma\right)\beta_0 + \beta_1 x_{i1} + \left(1-\gamma\right)\alpha_i - \gamma\epsilon_{i0} + \epsilon_{i1} \\ &= \beta_0^{\ddagger} + \beta_1^{\ddagger}X_{i1} + \epsilon_i^{\ddagger}.\end{aligned}$$

We can estimate $\gamma$ with Equation 15. Given $\hat{\gamma}$, the least squares estimate for $\beta^\ddagger$ is

$$\begin{pmatrix}\hat\beta_0^\ddagger\\ \hat\beta_1^\ddagger\end{pmatrix} \mid \hat\gamma = \left(X^\intercal X\right)^{-1} X^\intercal \left(Y_{:,1} - \hat\gamma Y_{:,0}\right) \tag{7}$$
$$= \begin{pmatrix}\left(1-\hat\gamma\right)\beta_0\\ \beta_1\end{pmatrix} + \left(X^\intercal X\right)^{-1} X^\intercal \left(\left(1-\hat\gamma\right)\alpha_i + \hat\gamma\epsilon_{i0} + \epsilon_{i1}\right)$$
$$\sim \mathcal{N}\left(\begin{pmatrix}\left(1-\hat\gamma\right)\beta_0\\ \beta_1\end{pmatrix}, \left(\left(1-\hat\gamma\right)^2\sigma_\alpha^2 + \hat\gamma^2\sigma_\epsilon^2 + \sigma_\epsilon^2\right)\left(X^\intercal X\right)^{-1}\right).$$

Regardless of $\hat\gamma$, $\hat\beta_1^\ddagger$ is an unbiased estimate of $\beta_1$, for

$$\mathbb{E}\left[\hat\beta_1^\ddagger\right] = \mathbb{E}_{\hat\gamma}\left[\mathbb{E}\left[\hat\beta_1^\ddagger \mid \hat\gamma\right]\right]\mathbb{E}_{\hat\gamma}\left[\beta_1\right] = \beta_1$$

by law of total expectation.

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

$$\mathbb{E}\left[\hat\beta_1^\dagger\right] = \mathbb{E}\left[\hat\beta_1^\star\right] = \mathbb{E}\left[\hat\beta_1^\ddagger\right] = \beta_1, \tag{8}$$

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 8 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

$$X^\intercal X = \begin{pmatrix}2n & \sum_{i=1}^{2n} x_{i1}\\ \sum_{i=1}^{2n} x_{i1} & \sum_{i=1}^{2n} x_{i1}^2\end{pmatrix} = \begin{pmatrix}2n & n\\ n & n\end{pmatrix}$$
$$\implies \left(X^\intercal X\right)^{-1} = \frac{1}{n^2}\begin{pmatrix}n & -n\\ -n & 2n\end{pmatrix} = \frac{1}{n}\begin{pmatrix}1 & -1\\ -1 & 2\end{pmatrix}. \tag{9}$$

Thus, we find that

$$\operatorname{var}\left(\hat\beta_1^\dagger\right) = \frac{2}{n}\left(\sigma_\alpha^2 + \sigma_\epsilon^2\right). \tag{10}$$

**$\hat\beta_1^\star$:** Using Equations 4 and 9, we compute that

$$\operatorname{var}\left(\hat\beta_1^\star\right) = \frac{4}{n}\sigma_\epsilon^2. \tag{11}$$

**$\hat\beta_1^\ddagger$:** We use Equation 7 to compute the variance conditional in terms of $\hat\gamma$. First, we note that

$$\operatorname{var}\left(\hat\beta_1^\ddagger\right) = \left(\left(1-\hat\gamma\right)^2\sigma_\alpha^2 + \hat\gamma^2\sigma_\epsilon^2 + \sigma_\epsilon^2\right)\left(X^\intercal X\right)^{-1}_{22}$$
$$= \frac{2}{n}\left(\left(1-\hat\gamma\right)^2\sigma_\alpha^2 + \hat\gamma^2\sigma_\epsilon^2 + \sigma_\epsilon^2\right)$$
$$= \frac{2}{n}\left(\hat\gamma^2\left(\sigma_\alpha^2 + \sigma_\epsilon^2\right) - 2\hat\gamma\sigma_\alpha^2 + \sigma_\alpha^2 + \sigma_\epsilon^2\right). \tag{12}$$

From Equations 10 and 11, whether the follow-up model or change model estimates $\beta_1$ more efficiently depends on whether the variance of the random effect is larger than the random effect of the measurement error. When the variance of the random effect is larger ($\sigma_\alpha^2 > \sigma_\epsilon^2$), $\mathrm{var}\left(\hat{\beta}_1^\star\right) < \mathrm{var}\left(\hat{\beta}_1^\dagger\right)$, so the change model is more efficient. Otherwise if $\sigma_\alpha^2 < \sigma_\epsilon^2$, the follow-up model is more efficient.

The ANCOVA model is more interesting. From Equation 12, when $\hat{\gamma} = 0$, efficiency is the same as the follow-up model, and when $\hat{\gamma} = 1$, efficiency is the same as the change model. $\mathrm{var}\left(\hat{\beta}_1^\ddagger\right)$ is a strictly convex function of $\hat{\gamma}$ which is minimized at

$$\hat{\gamma}^* = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\epsilon^2}. \tag{13}$$

By the Gauss-Markov theorem that states that the ordinary least squares estimate gives the lowest variance estimate for unbiased estimators, we must have that $\hat{\gamma} = \hat{\gamma}^*$, so

$$\begin{aligned}\mathrm{var}\left(\hat{\beta}_1^\ddagger\right) &= \frac{2}{n}\left((1-\hat{\gamma}^*)^2\,\sigma_\alpha^2 + (\hat{\gamma}^*)^2\,\sigma_\epsilon^2 + \sigma_\epsilon^2\right)\\ &= \frac{2}{n}\left(\frac{\sigma_\alpha^2\sigma_\epsilon^2}{\sigma_\alpha^2+\sigma_\epsilon^2} + \sigma_\epsilon^2\right) \lneq \frac{2}{n}\left(\max\left(\sigma_\alpha^2, \sigma_\epsilon^2\right) + \sigma_\epsilon^2\right),\end{aligned} \tag{14}$$

which results in $\hat{\beta}_1^\ddagger$ being a more efficient estimator than both $\hat{\beta}_1^\dagger$ and $\hat{\beta}_1^\star$. The behavior of $\hat{\gamma}$ will be investigated more fully in the next two parts.

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

**Solution:** From Equation 7, we have that

$$\begin{aligned}\hat{\gamma} &= \left(\left(\left(X^\ddagger\right)^\intercal X^\ddagger\right)^{-1}\left(X^\ddagger\right)^\intercal Y_{:,1}\right)_3 \\ &= \sum_{k=1}^{3}\left(\left(X^\ddagger\right)^\intercal X^\ddagger\right)^{-1}_{3k}\left(\left(X^\ddagger\right)^\intercal Y_{:,1}\right)_k \\ &= -n\frac{\sum_{i=1}^n Y_{i0}}{\det\left(\left(X^\ddagger\right)^\intercal X^\ddagger\right)}\sum_{i=1}^{2n} Y_{i1} + n\frac{\sum_{i=1}^n Y_{i0} - \sum_{i=n+1}^{2n} Y_{i0}}{\det\left(\left(X^\ddagger\right)^\intercal X^\ddagger\right)}\sum_{i=n+1}^{2n} Y_{i1} \\ &\quad + \frac{n^2}{\det\left(\left(X^\ddagger\right)^\intercal X^\ddagger\right)}\sum_{i=1}^{2n} Y_{i0}Y_{i1} \\ &= \frac{n}{\det\left(\left(X^\ddagger\right)^\intercal X^\ddagger\right)}\left(n\sum_{i=1}^{2n} Y_{i0}Y_{i1} - \sum_{i=1}^{n} Y_{i0}\sum_{i=1}^{n} Y_{i1} - \sum_{i=n+1}^{2n} Y_{i0}\sum_{i=n+1}^{2n} Y_{i1}\right),\end{aligned} \tag{15}$$

where $n \big/ \det\left(\left(X^\ddagger\right)^\intercal X^\ddagger\right)$ can be obtained from Equation 16. One can also write Equation 15 in terms of empirical variance estimates as in Equation 19.

$$
\begin{aligned}
\frac{1}{n}\det\left(\left(X^{\ddagger}\right)^{\intercal}X^{\ddagger}\right) &= n\sum_{i=1}^{2n}Y_{i0}^2-\left(\sum_{i=1}^{n}Y_{i0}\right)^2-\left(\sum_{i=1}^{n}Y_{i0}\right)\left(\sum_{i=n+1}^{2n}Y_{i0}\right)\\
&\quad-\left(\sum_{i=n+1}^{2n}Y_{i0}\right)\left(2\sum_{i=n+1}^{2n}Y_{i0}-\sum_{i=1}^{2n}Y_{i0}\right)\\
&= n\sum_{i=1}^{2n}Y_{i0}^2-\left(\sum_{i=1}^{n}Y_{i0}\right)^2-\left(\sum_{i=n+1}^{2n}Y_{i0}\right)^2. \qquad (16)
\end{aligned}
$$

(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

$$
\begin{aligned}
\widehat{\mathrm{cov}}\left(Y_{i0},Y_{i1}\mid x_{i1}=0\right) &= \frac{1}{n}\sum_{i=1}^{n}Y_{i0}Y_{i1}-\left(\frac{1}{n}\sum_{i=1}^{n}Y_{i0}\right)\left(\frac{1}{n}\sum_{i=1}^{n}Y_{i1}\right) \qquad (17)\\
\widehat{\mathrm{cov}}\left(Y_{i0},Y_{i1}\mid x_{i1}=1\right) &= \frac{1}{n}\sum_{i=n+1}^{2n}Y_{i0}Y_{i1}-\left(\frac{1}{n}\sum_{i=n+1}^{2n}Y_{i0}\right)\left(\frac{1}{n}\sum_{i=n+1}^{2n}Y_{i1}\right),
\end{aligned}
$$

respectively. Similarly, we can denote the MLE of the variances of $Y_{i0}$ without and with the intervention by

$$
\begin{aligned}
\widehat{\mathrm{var}}\left(Y_{i0}\mid x_{i1}=0\right) &= \frac{1}{n}\sum_{i=1}^{n}Y_{i0}^2-\left(\frac{1}{n}\sum_{i=1}^{n}Y_{i0}\right)^2 \qquad (18)\\
\widehat{\mathrm{var}}\left(Y_{i0}\mid x_{i1}=1\right) &= \frac{1}{n}\sum_{i=n+1}^{2n}Y_{i0}^2-\left(\frac{1}{n}\sum_{i=n+1}^{2n}Y_{i0}\right)^2,
\end{aligned}
$$

respectively. Substituting Equations 17 and 18 into the numerator and denominator of Equation 15, respectively, we can write

$$
\hat{\gamma}=\frac{\widehat{\mathrm{cov}}\left(Y_{i0},Y_{i1}\mid x_{i1}=0\right)+\widehat{\mathrm{cov}}\left(Y_{i0},Y_{i1}\mid x_{i1}=1\right)}{\widehat{\mathrm{var}}\left(Y_{i0}\mid x_{i1}=0\right)+\widehat{\mathrm{var}}\left(Y_{i0}\mid x_{i1}=1\right)}, \qquad (19)
$$

so we can interpret $\gamma$ as the overall autocorrelation between $Y_{i0}$ and $Y_{i1}$. Based on the true model, we can compute

$$
\begin{aligned}
\mathrm{var}\left(Y_{i0}\right) &= \mathrm{var}\left(\epsilon_{i0}\right)+\mathrm{var}\left(\alpha_i\right)=\sigma_\epsilon^2+\sigma_\alpha^2\\
\mathrm{cov}\left(Y_{i0},Y_{i1}\mid x_{i1}=0\right) &= \mathrm{cov}\left(Y_{i0},Y_{i1}\mid x_{i1}=1\right)\\
&= \mathbb{E}\left[\left(\alpha_i+\epsilon_{i0}\right)\left(\alpha_i+\epsilon_{i1}\right)\right]\\
&= \mathrm{var}\left(\alpha_i\right)=\sigma_\alpha^2,
\end{aligned}
$$

so the expected value of $\hat{\gamma}$ is

$$
\mathbb{E}\left[\hat{\gamma}\right]\approx\frac{\sigma_\alpha^2}{\sigma_\alpha^2+\sigma_\epsilon^2} \qquad (20)
$$

for large $n$ by Stutsky's theorem, which is the value in Equation 13 that minimizes the variance.