Midterm: STAT 570

Philip Pham

November 19, 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},

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

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

$$\begin{pmatrix} \hat{\beta}_0^{\ddagger} \\ \hat{\beta}_1^{\ddagger} \end{pmatrix} \mid \hat{\gamma} = \left(X^{\intercal}X\right)^{-1} X^{\intercal} \left(Y_{:,1} - \hat{\gamma} Y_{:,0}\right) \tag{7}$$
$$= \begin{pmatrix} (1-\hat{\gamma})\,\beta_0 \\ \beta_1 \end{pmatrix} + \left(X^{\intercal}X\right)^{-1} X^{\intercal} \left((1-\hat{\gamma})\,\alpha_i + \hat{\gamma}\epsilon_{i0} + \epsilon_{i1}\right)$$
$$\sim \mathcal{N}\left(\begin{pmatrix} (1-\hat{\gamma})\,\beta_0 \\ \beta_1 \end{pmatrix}, \left((1-\hat{\gamma})^2 \sigma_\alpha^2 + \hat{\gamma}^2\sigma_\epsilon^2 + \sigma_\epsilon^2\right)\left(X^{\intercal}X\right)^{-1}\right).$$

Regardless of $\hat{\gamma}$, $\hat{\beta}_1^{\ddagger}$ is an unbiased estimate of β_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

$$\operatorname{var}\left(\hat{\beta}_1^{\ddagger}\right) = \left((1-\hat{\gamma})^2 \sigma_\alpha^2 + \hat{\gamma}^2\sigma_\epsilon^2 + \sigma_\epsilon^2\right)\left(X^{\intercal}X\right)_{22}^{-1}$$
$$= \frac{2}{n}\left((1-\hat{\gamma})^2 \sigma_\alpha^2 + \hat{\gamma}^2\sigma_\epsilon^2 + \sigma_\epsilon^2\right)$$
$$= \frac{2}{n}\left(\hat{\gamma}^2\left(\sigma_\alpha^2 + \sigma_\epsilon^2\right) - 2\hat{\gamma}\sigma_\alpha^2 + \sigma_\alpha^2 + \sigma_\epsilon^2\right). \tag{12}$$

From Equations 10 and 11, whether the follow-up model or change model estimates β_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 ($\sigma_\alpha^2 > \sigma_\epsilon^2$), $\operatorname{var}\left(\hat{\beta}_1^\star\right) < \operatorname{var}\left(\hat{\beta}_1^\dagger\right)$, 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. $\operatorname{var}\left(\hat{\beta}_1^\ddagger\right)$ 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

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

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

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

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.

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

(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

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

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

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

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

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

so we can interpret γ 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

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

for large n by Stutsky's theorem, which is the value in Equation 13 that minimizes the variance, so Equation 19 agrees with our result in Equation 14. Indeed, results from Section 3 of Unbiased Estimation of Certain Correlation Coefficients tell us that $\mathbb{E}\left[\hat{\gamma}\right] = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\epsilon^2}$.

(e) Briefly discuss the implications on inference under each of the three models in the situation in which the intervention is non-randomized.

Solution: Suppose the intervention is non-randomized, that is, $\operatorname{cov}\left(Y_{i0}, x_{i1}\right) \neq 0$. This implies that at least one of $\operatorname{cov}\left(\alpha_i, x_{i1}\right) \neq 0$ or $\operatorname{cov}\left(\epsilon_{i0}, x_{i1}\right) \neq 0$ is true. The various error terms may no longer be centered

$$
\begin{aligned}
\epsilon_{i1}^\dagger &= \alpha_i + \epsilon_{i1} \\
\epsilon_{i1}^\star &= -\epsilon_{i0} + \epsilon_{i1} \\
\epsilon_{i1}^\ddagger &= \left(1 - \gamma\right)\alpha_i - \gamma\epsilon_{i0} + \epsilon_{i1}
\end{aligned}
$$

at 0 depending on the nature of the covariance which violates an assumption of the Gauss-Markov theorem, so we may no longer obtain an unbiased estimate of β_1.

Specifically, we would find

$$
\begin{aligned}
\mathbb{E}\left[\hat{\beta}_1^\dagger\right] &= \beta_1 + \left(\mathbb{E}\left[\epsilon_{i1}^\dagger \mid x_{i1} = 1\right] - \mathbb{E}\left[\epsilon_{i1}^\dagger \mid x_{i1} = 0\right]\right) \\
\mathbb{E}\left[\hat{\beta}_1^\star\right] &= \beta_1 + \left(\mathbb{E}\left[\epsilon_{i1}^\star \mid x_{i1} = 1\right] - \mathbb{E}\left[\epsilon_{i1}^\star \mid x_{i1} = 0\right]\right) \\
\mathbb{E}\left[\hat{\beta}_1^\ddagger\right] &= \beta_1 + \left(\mathbb{E}\left[\epsilon_{i1}^\ddagger \mid x_{i1} = 1\right] - \mathbb{E}\left[\epsilon_{i1}^\ddagger \mid x_{i1} = 0\right]\right).
\end{aligned}
$$

So, if x_{i1} is only correlated with ϵ_{i0}, $\hat{\beta}_1^\dagger$ will still be an unbiased estimator. If x_{i1} is only correlated with α_i, $\hat{\beta}_1^\star$ will still be unbiased. Since $\epsilon_{i1}^\ddagger$ is function of both ϵ_{i0} and α_i, $\hat{\beta}_1^\ddagger$ will no longer be an unbiased estimator.

2. Again in the context of child mortality, let

$$
\begin{aligned}
s_1 &= \mathbb{P}\left(\text{survived first year}\right) \\
s_2 &= \mathbb{P}\left(\text{survived years 1–5} \mid \text{survived first year}\right) \\
s_3 &= \mathbb{P}\left(\text{survived first 5 years}\right) = s_1 \times s_2. \qquad (21)
\end{aligned}
$$

Let $\hat{s}_{1i}$, $\hat{s}_{2i}$, and $\hat{s}_{3i}$ be estimates of s_{1i}, s_{2i}, and s_{3i} for $i = 1, \ldots, n$ areas in a country. Let $Y_i = \log \hat{s}_{1i}$, $Z_i = \log \hat{s}_{2i}$, and $X_i = \log \hat{s}_{3i}$.

Let

$$
\begin{aligned}
\mathbb{E}\left[Y_i\right] &= \mu_1 \\
\operatorname{var}\left(Y_i\right) &= \Sigma_{11} \\
\mathbb{E}\left[Z_i\right] &= \mu_2 \\
\operatorname{var}\left(Z_i\right) &= \Sigma_{22} \\
\operatorname{cov}\left(Y_i, Z_i\right) &= \Sigma_{12}. \qquad (22)
\end{aligned}
$$

(a) Suppose Y_i and X_i have a bivariate normal distribution. Write down the mean vector and variance-covariance matrix for Y_i and X_i.

Solution: Let U_i and V_i be independent and identically distributed standard normal random variables. Then, we must have that

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} = \mu + A \begin{pmatrix} U_i \\ V_i \end{pmatrix}, \tag{23}$$

where $AA^\intercal = \Sigma$ if $\begin{pmatrix} X_i & Y_i \end{pmatrix}^\intercal$ is drawn from a bivariate normal. From Equation 22, we can choose

$$A = \begin{pmatrix} \sqrt{\Sigma_{11}} & 0 \\ \frac{\Sigma_{12}}{\sqrt{\Sigma_{11}}} & \sqrt{\frac{\Sigma_{11}\Sigma_{22} - \Sigma_{12}^2}{\Sigma_{11}}} \end{pmatrix}. \tag{24}$$

The parameterization must be

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix} \right). \tag{25}$$

(b) Show that $\mathbb{E}\left[Y_i \mid X_i = x\right]$ takes the form of a simple linear regression model and identify β_0 and β_1. Carefully interpret β_1.

Solution: From Equation 23, we have that

$$\begin{aligned} X_i &= \mu_1 + A_{11}U_i + A_{12}V_i \\ Y_i &= \mu_2 + A_{21}U_i + A_{22}V_i. \end{aligned}$$

If we fix $X_i = x$ and use the defintion of A in Equation 24, we have that

$$\begin{aligned} Y_i &= \mu_2 + A_{21}U + \frac{A_{22}}{A_{12}}\left(x - A_{11}U - \mu_1\right) \\ &= \mu_2 + \frac{A_{21}}{A_{11}}U + A_{22}V \\ &= \mu_2 + \frac{\Sigma_{12}}{\Sigma_{22}}\left(x - \mu_1\right) + \sqrt{\frac{\Sigma_{11}\Sigma_{22} - \Sigma_{12}^2}{\Sigma_{11}}}V \\ &\sim \mathcal{N}\left(\mu_2 + \frac{\Sigma_{12}}{\Sigma_{11}}\left(x - \mu_1\right), \frac{\Sigma_{11}\Sigma_{22} - \Sigma_{12}^2}{\Sigma_{11}}\right). \end{aligned} \tag{26}$$

Equation 26 gives us

$$\mathbb{E}\left[Y_i \mid X_i = x\right] = \mu_2 + \frac{\Sigma_{12}}{\Sigma_{11}}\left(x - \mu_1\right) = \mu_2 - \frac{\Sigma_{12}}{\Sigma_{11}}\mu_1 + \frac{\Sigma_{12}}{\Sigma_{11}}x. \tag{27}$$

Imagine fitting the model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i. \tag{28}$$

Equations 26 and 27 show that the ordinary least squares estimator for Equation 28 is

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \mid X \sim \mathcal{N}\left(\begin{pmatrix} \mu_2 - \frac{\Sigma_{12}}{\Sigma_{11}}\mu_1 \\ \frac{\Sigma_{12}}{\Sigma_{11}} \end{pmatrix}, \frac{\Sigma_{11}\Sigma_{22} - \Sigma_{12}^2}{\Sigma_{11}}\left(X^\intercal X\right)^{-1} \right), \tag{29}$$

	estimate	standard error	t-statistic	p-value
$\hat{\beta}_0$	-0.002178	0.002584	-0.842783	4.038069e-01
$\hat{\beta}_1$	0.686650	0.030340	22.631990	3.179129e-26

Table 1: Least squares estimates for the model in Equation 28.

where X is a $n \times 2$ matrix with 1s in the first column and $X_{i2} = X_i$ in the second column.

From Equations 28 and 29, we see that β_1 quantifies the correlation between X_i and Y_i: given an observation X_i that is 1 unit greater, we would expect to observe a difference of β_1 units in Y_i. Note that

$$\mathbb{E}\left[\hat{\beta}_1\right] = \frac{\Sigma_{12}}{\Sigma_{11}} = \rho\sqrt{\frac{\Sigma_{22}}{\Sigma_{11}}}, \tag{30}$$

where $\rho = \frac{\Sigma_{12}}{\sqrt{\Sigma_{11}\Sigma_{22}}}$ is the Pearson correlation coefficient, so β_1 is the correlation between X_i and Y_i scaled by the standard errors.

Rewriting the model in Equation 28 in terms of s_{1i} and s_{2i}, we have that

$$\begin{aligned} \log s_{1i} &= \beta_0 + \beta_1 \log s_{3i} + \epsilon_i \\ s_{1i} &= \exp(\beta_0) \times s_{3i}^{\beta_1} \times \exp(\epsilon_i) \\ s_{1i}^{1/\beta_1} = \left(\frac{1}{s_{1i}}\right)^{-1/\beta_1} &= \exp\left(\frac{\beta_0}{\beta_1}\right) \times s_{3i} \times \exp\left(\frac{\epsilon_i}{\beta_1}\right). \end{aligned}$$

β_1 indicates how much to weight the probability of having survived the first 5 years in computing the expected probability of having survived the first year. If one interprets the probability of dying as exponential process, $1/\beta_1$ can be seen as how fast the probability of survival decays over a 5 year period. Specifically, with each year the probability of suriviving decays $\frac{1}{5\beta_1}\log s_{1i}$ on average.

Another way to interpret this is by writing Equation 28 as

$$s_{1i} = \exp(\beta_0) \times s_{3i}^{\beta_1 - 1} \times s_{3i} \times \exp(\epsilon_i). \tag{31}$$

Then, $\exp(\beta_0) \times s_{3i}^{\beta_1 - 1}$ becomes an estimate of s_{2i}^{-1} since $s_{3i} = s_{1i} \times s_{2i}$.

Note that $s_{3i} \leq 1$, so $X_i \leq 0$. So large values of β_1 drive s_{i1} to 0. 0 would mean that X_i has no association with Y_i.

(c) The data on the website contain estimates of the log first year survival (Y_i) and log five year survival (X_i) for $n = 47$ areas in Kenya in the period 2000–2004. Fit a linear model and hence summarize the association between first year and first five year survival, in these data.

Solution: The result of fitting a linear model can be found in Table 1. The association of Y_i with X_i is statistically significant.

Over 5 years, the probability of survival decayed at a rate of $\frac{1}{5\beta_1}\log s_{1i} \approx 0.291 \log s_{1i}$ each year on average.

(d) Now suppose we are presented with a new area, for which only five year survival is available and known to be 0.95. Obtain point estimates and a 95% confidence interval for:

	Estimate	95% CI lower bound	95% CI upper bound
Surviving the first year	0.963292	0.946724	0.980150
Death within the first year	0.036708	0.019850	0.053276
Death between ages 1 and 5, given survival unti...	0.986201	0.969239	1.003460
Death between ages 1 and 5	0.013292	-0.003276	0.030150

Table 2: Predicted parameters of interest for a new area with $s_{3,0} = 0.95$.

i. Surviving the first year.
ii. Death within the first year.
iii. Death between ages 1 and 5, given survival until age 1.
iv. Death between ages 1 and 5.

Solution: Note that $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$, so a prediction at a given point x_0 is

$$\hat{y}_0 = \left(\bar{Y} - \hat{\beta}_1 \bar{X}\right) + \hat{\beta}_1 x_0 + \epsilon_0 = \bar{Y} + \hat{\beta}_1 \left(x_0 - \bar{X}\right) + \epsilon_0. \tag{32}$$

Let $\sigma^2 = \frac{\Sigma_{11}\Sigma_{22} - \Sigma_{12}^2}{\Sigma_{11}}$. We can estimate it with the MLE:

$$\hat{\sigma}^2 = \frac{\hat{\Sigma}_{11}\hat{\Sigma}_{22} - \hat{\Sigma}_{12}^2}{\hat{\Sigma}_{11}}, \tag{33}$$

where $\hat{\Sigma}$ is the sample covariance matrix for the X_i and Y_i.
For the variance of Equation 32, we have that

$$\operatorname{var}\left(\hat{y}_0 \mid x_0, X\right) = \sigma^2 + \frac{\sigma^2}{n} + \sigma^2 \left(X^\intercal X\right)^{-1}_{22} \left(x - \bar{X}\right)^2.$$

Since the sample size is fairly large, using the plug-in estimators, we have an approximate normal distribution,

$$\hat{y}_0 \mid x_0, X \sim \mathcal{N}\left(\hat{\beta}_0 + \hat{\beta}_1 x_0, \hat{\sigma}^2 \left(1 + \frac{1}{n} + \left(X^\intercal X\right)^{-1}_{22} \left(x - \bar{X}\right)^2\right)\right). \tag{34}$$

Let Φ be the cumulative distribution function of the normal distribution. The 95% confidence interval is

$$\left[\hat{y}_0 + \Phi^{-1}(0.025)\sqrt{\hat{\operatorname{var}}\left(\hat{y}_0 \mid x_0, X\right)}, \hat{y}_0 + \Phi^{-1}(0.975)\sqrt{\hat{\operatorname{var}}\left(\hat{y}_0 \mid x_0, X\right)}\right],$$

which we can compute with Equation 34. $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$ The true log probability of surviving 1 year in the new area with should fall in this compute confidence interval 95% of the time this study is done.
To estimate the quantities of interest, we can reparameterize. The probability of surviving the first year is $\exp\left(\hat{y}_0\right)$. The probability of death is $1 - \exp\left(\hat{y}_0\right)$. Since $s_{2i} = s_{3i}/s_{1i}$, we have that the estimated probability of survival having survived the first year is $\exp\left(x_0\right) / \exp\left(\hat{y}_0\right)$. Death between the ages of 1 and 5 is $s_{1i} \times (1 - s_{2i})$, which can be estimated as

$$\exp\left(\hat{y}_0\right) \times \left(1 - \frac{\exp\left(x_0\right)}{\exp\left(\hat{y}_0\right)}\right) = \exp\left(\hat{y}_0\right) - \exp\left(x_0\right).$$

Since these are all monotonic transformations of $\hat{y}_0$ confidence intervals can be computed by reparameterizing the end points. These estimates can be seen in Table 2 for $x_0 = \log(0.95)$.

Computations can be found in `mortality.ipynb`.

(e) Can you see any problems with this model?

Solution: Yes. Recall from the problem description that $\exp(X_i) = s_{3i} = s_{1i} \times s_{2i} = \exp(Y_i)\exp(Z_i) = \exp(Y_i + Z_i)$. The probability of surviving 1 year must be greater than or equal to the probability of surviving 5 years, but the model doesn't enforce this constraint, so inference can produce strange results.

In particular, the probabilities s_{1i}, s_{2i}, and s_{3i} are always between 0 and 1, so Y_i and X_i cannot possibly be normally distributed. We see how this is problematic in Table 2 where we obtain probabilites that are less than 0 or greater than 1. Negative values of β_1 could produce estimates of s_{1i} greater than 1, too.

More subjectively, the model is hard to interpret. It's more natural to think about the probability of surviving 5 years given the probability of having survived 1 year given the real-world generative model that one must have had to survived 1 year in order to have survived 5 years.