

Problem 1

- (a) The study does not appear to be randomized. While the group means are the same at Y_{i1} , the group variances at Y_{i1} are not the same. Randomization would mean the group means and variances are the same at baseline.

(b)

$$\begin{aligned}
 E(D_i; x_i) &= \alpha_1 + \alpha_2 x_i \\
 E(D_i | x_i = 0) &= \alpha_1 = -10 \\
 E(D_i | x_i = 1) &= \alpha_1 + \alpha_2 = -20 \\
 \alpha_2 &= -20 - (-10) = -10 \\
 \alpha_1 &= -10 \quad \alpha_2 = -10 \\
 Var(D_i | x_i = 0) &= 20 + 30 - 2 * 10 = 30 \\
 Var(D_i | x_i = 1) &= 40 + 30 - 2 * 20 = 30
 \end{aligned}$$

- (c) α_2 is the difference between the group mean differences from baseline to time 2

(d)

$$\begin{aligned}
 E(D_i | Y_{i1}; x_i) &= E(Y_{i2} - Y_{i1} | Y_{i1}; x_i) \\
 &= E(Y_{i2} | Y_{i1}; x_i) - Y_{i1} \\
 Y_{i2} | Y_{i1} &\sim N(u_2 + \frac{\sigma_{12}}{\sigma_{11}}(Y_1 - \mu_1), \Sigma^*) \\
 E(Y_{i2} | Y_{i1}; x_i) - Y_{i1} &= u_2 + \frac{\sigma_{12}}{\sigma_{11}}(Y_1 - \mu_1) - Y_{i1} \text{ (for each group separately)} \\
 E(D_i | Y_{i1}; x_i = 0) &= 40 + \frac{10}{20}(Y_{i1} - 50) - Y_{i1} = 15 - .5Y_{i1} \\
 E(D_i | Y_{i1}; x_i = 1) &= 30 + \frac{20}{40}(Y_{i1} - 50) - Y_{i1} = 5 - .5Y_{i1}
 \end{aligned}$$

- (e) $E(D_i | Y_{i1}; x_i) = (15 - Y_{i1}) + [(5 - 15) + (-.5 + .5)Y_{i1}]x_i$
 $= 15 - .5Y_{i1} - 10x_i$

(f)

$$\begin{aligned}
E(Y_{i2}|Y_{i1}; x_i) &= \mu_2 + \frac{\sigma_{12}}{\sigma_{11}}(Y_{i1} - \mu_1) \text{ (for each group separately)} \\
E(Y_{i2}|Y_{i1}; x_i = 0) &= 40 + .5(Y_{i1} - 50) = 15 + .5Y_{i1} \\
E(Y_{i2}|Y_{i1}; x_i = 1) &= 30 + .5(Y_{i1} - 50) = 5 + .5Y_{i1} \\
E(Y_{i2}|Y_{i1}; x_i) &= (15 + .5Y_{i1}) + [(5 - 15) + (.5 - .5)Y_{i1}]x_i \\
&= 15 + .5Y_{i1} - 10x_i
\end{aligned}$$

$$\begin{aligned}
\text{(g) } \text{Var}(D_i|Y_{i1}; x_i) &= \text{Var}(Y_{i2} - Y_{i1}|Y_{i1}; x_i) \\
&= \text{Var}(Y_{i2}|Y_{i1}; x_i) + \text{Var}(Y_{i1}|Y_{i1}; x_i) - 2\text{Cov}([Y_{i2}|Y_{i1}; x_i], Y_{i1}|Y_{i1}; x_i]) \\
&= \text{Var}(Y_{i2}|Y_{i1}; x_i) + 0 + 0 \\
\text{Thus } \text{Var}(Y_{i2}|Y_{i1}; x_i) &= \text{Var}(D_i|Y_{i1}; x_i)
\end{aligned}$$

(h) Since the interaction term $x_i * Y_{i1}$ is 0, the assumption of parallel lines holds

(i) The OLS estimate of β_2 will be an unbiased estimate of α_2 since the group means are the same at baseline.

(j)

$$\begin{aligned}
\text{Var}(D_i|Y_{i1}; x_i) &= \text{Var}(Y_{i2}|Y_{i1}; x_i) \\
Y_{i2}|Y_{i1}; x_i &\sim N(\mu^*, \sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}) \\
\text{Var}(Y_{i2}|Y_{i1}; x_i = 0) &= 30 - \frac{10^2}{20} = 25 \\
\text{Var}(Y_{i2}|Y_{i1}; x_i = 1) &= 30 - \frac{20^2}{40} = 20
\end{aligned}$$

(k)

$$\begin{aligned}
&\text{using } n_0 = n_1 = 1 \\
\text{Var}(D_i|x_i = 0) &= 30 \\
\text{Var}(D_i|x_i = 1) &= 30 \\
\text{Variance not adjusting for baseline} &= V_0/1 + V_1/1 = 60 \\
\text{Var}(D_i|Y_i; x_i = 0) &= 25 \\
\text{Var}(D_i|Y_i; x_i = 1) &= 20
\end{aligned}$$

Variance adjusting for baseline = $V_0/1 + V_1/1 = 45$

Efficiency Ratio = Var not adj for baseline / Var adj baseline = $60/45 = 1.333$

There is a 33.33% gain in efficiency brought about by adjusting for baseline

Problem 2

The sample size is 1000 as defined by the variable n

A mean column vector μ is created of length 4

A covariance matrix, Σ is created which is 4x4 and it does not have homogeneous variance. Σ is symmetric and positive definite.

Cholesky decomposition is performed on the covariance matrix, resulting in the upper Cholesky root of the covariance matrix, $U_{4 \times 4}$ an upper triangular matrix, with each value rounded to the nearest tenth. Since Σ is positive definite, all values of U will be positive.

The transpose of U is the lower Cholesky root and thus multiplying $U^T U$ results a matrix equivalent to the covariance matrix Σ .

A simulation is then run creating a random seed matrix which is 1000×4

A random standard normal variable is used to generate a random matrix and is multiplied by U, resulting in a matrix $Z_{1000 \times 4}$

A sample response matrix $y_{1000 \times 4}$ is simulated by adding z and a matrix of 1000 repeated rows of the transpose of the mean vector. The dimensions of the repeated mean matrix are 1000×4

\bar{y} , sample mean column vector of length 4 is created by taking the transpose of the mean of the simulated sample response matrix y.

The mean vector mu and the sample mean vector \bar{y} are printed side by side and the values are similar.

A sample covariance matrix $s_{4 \times 4}$ is created by taking the covariance of the simulated sample response matrix.

The covariance matrix and the simulated sample covariance matrix are printed and the sample covariance matrix slight larger values for the variances. The covariances in the sample covariance matrix are larger than in Σ . This difference is more pronounced than that of the variances.