

problem set 2

Problem 1 - Chapter 6 (Stats Mods), Exercise set C, 2-4 (p.90)

2. We need to compute the path coefficients in figure 2. If we look at Freedman p. 84, we see that we can calculate these coefficients using the standard OLS regression coefficient algorithm: $(M'M)^{-1}M'U$, utilizing the fact that the variables of interest are standardized. Particularly,

$$M'M = n \begin{bmatrix} 1 & r_{VX} \\ r_{VX} & 1 \end{bmatrix} = n \begin{bmatrix} 1 & 0.52 \\ 0.52 & 1 \end{bmatrix}$$

and

$$M'U = n \begin{bmatrix} r_{VU} \\ r_{XU} \end{bmatrix} = n \begin{bmatrix} -0.26 \\ -0.42 \end{bmatrix}$$

So

$$(M'M)^{-1} = \frac{1}{n(1 - (0.52)^2)} \begin{bmatrix} 1 & -0.52 \\ -0.52 & 1 \end{bmatrix}$$

and further

$$\begin{aligned} (M'M)^{-1}M'U &= \frac{1}{n(1 - (0.52)^2)} \begin{bmatrix} 1 & -0.52 \\ -0.52 & 1 \end{bmatrix} \times n \begin{bmatrix} -0.26 \\ -0.42 \end{bmatrix} \\ &= \begin{bmatrix} -0.06 \\ -0.39 \end{bmatrix} \end{aligned}$$

3. We would like to estimate the standard deviation of δ in equation 10. For this, we can turn to Freedman p.85. That is, we can plug in our values into Freedman's equation (8) and solve for $\hat{\sigma}^2$, then taking the square root to get the SD. We do this here. Thus we get 0.9055186 as the square root. Yet, Freedman adds a caveat that with small samples this isn't a good way to estimate sigma squared since degrees of freedom aren't taken into account. Thus in order to take them into account we define $p = 3$ and we know $n = 36$ and divide $\hat{\sigma}^2$ by $n/(n - p)$, thus divide 0.819964 by $36/(36 - 3) = 36/33$. We find $\hat{\sigma}^2$ to be equal to 0.7516337 after this adjustment. We take the square root to get the standard deviation at 0.8669681.

$$\begin{aligned} 1 &= \hat{a}^2 + \hat{b}^2 + 2\hat{a}\hat{b}r_{VX} + \hat{\sigma}^2 \\ 1 &= -0.06^2 + -0.39^2 + 2(-0.06 \times -0.39 \times 0.52) + \hat{\sigma}^2 \\ \hat{\sigma}^2 &= 1 - (-0.06^2 + -0.39^2 + 2(-0.06 \times -0.39 \times 0.52)) \\ \hat{\sigma}^2 &= 1 - (0.0036 + 0.1521 + 0.024) \\ \hat{\sigma}^2 &= 0.819964 \\ \hat{\sigma} &= 0.9055186 \end{aligned}$$

4. Find standard errors for path coefficients and their difference; and also t-ratios and statistical significance. Recall we need to estimate the variance covariance matrix - $\sigma^2(X'X)^{-1}$ and we have already estimated $\hat{\sigma}^2 = 0.7516337$. In the above exercise, we have also found

$$(X'X)^{-1} = \frac{1}{n(1 - (0.52)^2)} \begin{bmatrix} 1 & -0.52 \\ -0.52 & 1 \end{bmatrix}$$

So then

$$\hat{\sigma}^2(X'X)^{-1} = \frac{0.7516337}{36(1 - (0.52)^2)} \begin{bmatrix} 1 & -0.52 \\ -0.52 & 1 \end{bmatrix} = \begin{bmatrix} 0.0286 & -0.0148 \\ -0.0148 & 0.0286 \end{bmatrix}$$

And then we take the square roots of the diagonal in order to get the standard errors.

So the standard error of the path coefficient for mass tolerance ($\hat{\beta}_1 = -0.06$) is $\sqrt{0.0286} = 0.169$ and the standard error of the path coefficient for elite tolerance ($\hat{\beta}_2 = -0.39$) is $\sqrt{0.0286} = 0.169$. The standard error of the difference of the coefficients is the square root of the sum of the variances. Thus it is $\sqrt{0.0286 + 0.0286} = 0.2391652$

The t-ratio for the path coefficient for mass tolerance ($\hat{\beta}_1 = -0.06$) is $-0.06/0.169 = -0.355$, the t-ratio for the path coefficient for elite tolerance ($\hat{\beta}_2 = -0.39$) is $-0.39/0.169 = -2.307$ and the t-ratio for their difference is $-0.06 - -0.39/0.2391652 = 1.379$. The only statistically significant result at the 5 percent level is the coefficient for elite tolerance.

Problem 7

- total effect is c which we break down into direct and indirect effects. d is direct effect, ab is the indirect effect. the overarching derivation comes from the book (look at slides for this)
- why would this break down if it varies across subjects? there's a trick. we can re-arrange covariance as the expectation of the product of those things - product of the expectations. solve for e of xy - > ; then there exists at least one pair for which i and j are not the same. we can't compute the covariance. expectation of a and b we can't compute if we don't have the covariance - it breaks down in the indirect effect. aka its a very strong assumption to make to say that the indirect effect is constant.
- look at the mediator equation. if treatment effect is 0 for everyone - a is 0 (for the first equation.) $E[cX] = cE[X]$; $E[aib_i] = aE[b_i] = 0 * e[b_1] = 0$; when we deal with averages, we might have to use expectations.
- why does the complex potential outcome defy empirical investigation? you can never observe it. you can't at the same time set and not set z to 1.
- y1 given d = 1 - y0 given d = 1; one of these terms i dont observe.the distinction is that the in this equation 2 things are moving (z 0/1 and other one 0/1) whereas in the other, we fix z and manipulate the mediator.