Ec141, Spring 2019

Midterm 1 (alt)

Please read each question carefully. Start each question on a new bluebook page. The use of calculators and other computational aides is not allowed. Good luck!

[1] **[10 Points]** Please write your full name on this exam sheet and turn it in with your bluebook.

[2] **[30 Points]** Let $C_t = 1$ if an individual (child) went to college and zero otherwise. Let $C_{t-1} = 1$ if the corresponding individual's parent went to college and zero otherwise. The following table gives the joint distribution of father and sons' college attendance:

	$C_t = 0$	$C_t = 1$
$C_{t-1} = 0$	0.60	0.20
$C_{t-1} = 1$	0.10	0.10

For example 20% percent of the population consists of pairs with a father who did not attend college, but a son who did.

[a] **[5 Points]** Among children of college graduates, what fraction go on to complete college themselves? Among children of non-graduates, what fraction go on to complete college themselves?

[b] **[10 Points]** Let $\mathbb{E}^*\left[C_t|\, C_{t-1}\right] = a + bC_{t-1}$; calculate a and b.

[c] **[5 Points]** The following table gives child's adult earnings, Y_t, for each of the four subpopulations introduced above

	$C_t = 0$	$C_t = 1$
$C_{t-1} = 0$	\$8,000	\$60,000
$C_{t-1} = 1$	\$14,000	\$30,000

What is the average earnings level of college graduates in this economy? What is the average earnings of non-college graduates? What is the overall average earnings level? Express your answers symbolically using the notation of (conditional) expectations and also provide a numerical answer.

[d] **[5 Points]** Let $\pi_{c_{t-1}} = \Pr\left(C_{t-1} = c_{t-1}|\, C_t = 1\right)$. Consider the estimand

$$\beta = \sum_{c_{t-1}=0,1} \left\{\mathbb{E}\left[Y|\, C_t = 1, C_{t-1} = c_{t-1}\right] - \mathbb{E}\left[Y|\, C_t = 0, C_{t-1} = c_{t-1}\right]\right\} \pi_{c_{t-1}}.$$

In what sense does β adjust for "covariate differences" between college and non-college graduates? Evaluate β and compare your numerical answer with the raw college - non-college earnings gap you calculated in part (c). Why are these two numbers different?

[e] **[5 Points]** Gavin Newsom is considering a community college expansion policy. You have been tasked to asked to predict the earnings gain associated with completing a college degree. Gavin estimates that after the community college expansion the distribution of college attendance in California will look like

	$C_t = 0$	$C_t = 1$
$C_{t-1} = 0$	0.40	0.40
$C_{t-1} = 1$	0.05	0.15

Calculate average earnings in this new economy (you may assume that the mapping from background and education into earnings introduced in part (c) remains the same)?

[3] **[10 Points]** The Undergraduate Dean has been collecting data on the high school GPA (X) of incoming students for a long, long time. She has also kept track of 1st semester GPA (Y) for incoming students over the same period of time. She would like to be able to predict 1st semester GPA for incoming students using their high school GPA. She reports to you the following means, variances and a covariance for X and Y:

$$\mu_X = \frac{12}{5}, \mu_Y = 2$$

and

$$\sigma_X^2 = 1/6, \sigma_Y^2 = 1/6, \sigma_{XY} = 1/5.$$

Because she has collected such a large sample you are free to treat these numbers as if they were population quantities.

[a] **[5 Points]** Calculate the α and β associated with the (mean square error minimizing) linear predictor of Y given X, $\mathbb{E}^*[Y|X] = \alpha + \beta X$?

[b] **[5 Points]** What is the coefficient of determination associated with $\mathbb{E}^*[Y|X]$?

[4] **[20 Points]** The World Health Organization has contracted you to design a randomized experiment evaluating the efficacy of zinc supplements on diarrhea prevalence (measured as the number of episodes in the one hundred days prior to surveying). Let $Y(1)$ be the potential number of episodes of diarrhea if taking zinc supplements and $Y(0)$ the control potential outcome. A baseline survey of your target population yields a diarrhea prevalence of 10 days per one hundred days with a standard deviation of 5 days. Let N be your target sample size and assume that half of respondents will be randomly assigned to treatment. Assume that the variance of $Y(1)$ and $Y(0)$ are equal to each other. Also assume that no respondents in your baseline survey were taking zinc supplements.

[a] **[10 Points]** Derive an expression for the ex ante probability (β) that you reject the null of no effect in favor of a *one-sided* alternative of a negative effect (i.e., treatment reduces diarrhea). Let α denote the size of your test and θ the ATE. Carefully explain your reasoning and notation **[4-6 sentences]**.

[b] **[5 Points]** Assume that $\theta = -2$. How large would N need to be to ensure an ex ante rejection probability of 95 percent (for a test with size $\alpha = 0.05$).

[c] **[5 Points]** You ultimately design an experiment with power of $\beta = 0.95$ and size $\alpha = 0.10$. In the end you find no effect of zinc supplements on the prevalence of diarrhea (i.e., you fail to reject the null of no effect). Prior to the experiment you believed that the probability that zinc supplements reduced the prevalence of diarrhea was 0.75. What is your belief after your null finding? Explain **[2-4 sentences]**.

[5] **[30 Points]** Let W, X be a pair of regressors with the property that $\mathbb{C}(W, X) = 0$. Show that, for outcome, Y,

$$\mathbb{E}^*[Y|W, X] = \mathbb{E}^*[Y|W] + \mathbb{E}^*[Y|X] - \mathbb{E}[Y]. \tag{1}$$

You may assume that all objects in the above expression are well-defined (i.e., all necessary moments exist and so on).

[a] **[10 Points]** First show that

$$\mathbb{E}^*[\mathbb{E}^*[Y|W]|X] = \mathbb{E}^*[\mathbb{E}^*[Y|X]|W] = \mathbb{E}[Y]$$

[b] **[10 Points]** Second verify (1) using the Projection Theorem.

[c] **[10 Points]** Finally show that

$$\mathbb{E}^* [Y | W, X] = \mathbb{E} [Y] + \frac{\mathbb{C} (Y, W)}{\mathbb{V} (W)} (W - \mathbb{E} [W]) + \frac{\mathbb{C} (Y, X)}{\mathbb{V} (X)} (X - \mathbb{E} [X]) .$$