Ec141, Spring 2020

Midterm 1

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] **[5 Points]** Please write your full name and student ID on the first page of your pdf file of exam solutions.

[2] **[15 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 enrolled students over the same period of time. She would like to be able to predict a student's 1st semester GPA 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

$$\sigma_X^2 = 1/6, \sigma_Y^2 = 1/4, \sigma_{XY} = 1/5.$$

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] **[10 Points]** Say $\sigma_X^2 = 0$ (e.g., only students with 4.0 GPAs have ever been admitted). What is the best linear predictor of Y given X in this case (you may assume all other parameters stay the same)? Why? **[2-4 sentences]**.

[3] **[25 Points]** Consider the following statistical model for the logarithm of daily city-wide sales of Bob Dylan's landmark *Christmas in the Heart* album:

$$\ln S = \alpha_0 + \beta_0 R + \gamma_0 P + U, \ \mathbb{E}[U|\, R, P] = 0,$$

where R is the number of times a song from the album is played on KALX on the given day, and P is the price of the album (which varies across your sample due to various (exogenous) record label promotions, holiday sales and so on). A friend estimates $\theta_0 = (\alpha_0, \beta_0, \gamma_0)'$ by the method of least squares. She claims that $\sqrt{N}\left(\hat{\theta} - \theta_0\right) \xrightarrow{D} \mathcal{N}(0, \Lambda_0)$ and reports the following:

$$\hat{\theta} = \begin{pmatrix} 1.0 \\ 0.01 \\ -0.51 \end{pmatrix}, \ \frac{\hat{\Lambda}}{N} = \begin{pmatrix} 0.25 & -0.002 & 0.010 \\ -0.002 & 0.01 & 0.005 \\ 0.010 & 0.005 & 0.03 \end{pmatrix}.$$

[a] **[2 Points]** Calculate a 95 confidence interval for β_0.

[b] **[5 Points]** Your friend would like to test the hypothesis that "for Bob Dylan one song on the radio is as good as cutting record price by \$1" (a phrase used by her record store boss). Explain why this corresponds

to:

$$H_0 : \beta_0 = -\gamma_0$$
$$H_1 : \beta_0 \neq -\gamma_0$$

[c] **[5 Points]** We can re-write H_0 as

$$H_0 : C\theta = c$$

Provide the appropriate forms for C and c.

[d] **[5 Points]** How many restrictions on θ does H_0 imposes?

[e] **[5 Points]** Calculate the Wald statistics for H_0. Can we reject with size $\alpha = 0.05$?

[f] **[8 Points]** Now formalize and test the hypothesis that "for Bob Dylan one song on the radio is as good as cutting record price by \$3".

[4] **[30 Points]** Consider the following statistical model for the earnings of Berkeley students

$$Y = \alpha + \beta G + \gamma A + U,\ \mathbb{E}[U|G, A] = 0,$$

where G equals one if the student graduated and zero if they dropped out and A equals one if at least one of the student's parents graduated from college and zero otherwise.

[a] **[5 Points]** You read in the Oakland Tribune newspaper that Berkeley graduates earn an average of \$75,000 per year nationwide, while the earnings of dropouts average only \$15,000. Express this population earnings difference between Berkeley graduates and dropouts in terms of the statistical model given above.

[b] **[5 Points]** Under what conditions is it true that $\beta = \$60,000$? Do you think these conditions are likely to be true in practice? Briefly explain your answer **[3-5 sentences]**.

[c] **[5 Points]** The same article reports that among Berkeley graduates, three fourths come from families where at least one parent completed college, while among all former students (i.e., graduates and dropouts) only seven twelfths come from such families. It also states that the overall (i.e., unconditional) graduation rate at Berkeley is two-thirds. Among dropouts, what fraction come from families where at least one parent completed college?

[d] **[5 Points]** Assume $\gamma = \$25,000$. Using your answers in parts (a) and (c) solve for β. What is the expected earnings gain associated with graduating from Berkeley holding parent's education (i.e., A) constant? Briefly comment on why your answer differs from the earnings gap between graduates and dropouts reported by the Tribune **[3-5 sentences]**.

[e] **[5 Points]** You are considering dropping out of Cal to spend more time on Telegraph Avenue. What is the (approximate) expected earnings loss associated with this decision? Explain **[3-5 sentences]**.

[f] **[5 Points]** You move to Oakland upon graduation, your neighbor to the left tells you that he dropped out of Berkeley during the Free Speech Movement, your neighbor to the right tells you that he graduated from Berkeley about the same time. What is your expectation of the annual earnings of your two neighbors? Explain **[3-5 sentences]**.

[5] **[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 $\theta = -10$. 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 $\beta = 0.90$ and size $\alpha = 0.05$. 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 **[3-5 sentences]**.