Ec141, Spring 2020

Professor Bryan Graham

Review Sheet 1

This review sheet is designed to assist you in your exam preparations. I suggest preparing written answers to each question. You may find it useful to study with your classmates (indeed I encourage you to do so and also to be generous with one another as you prepare). In the exam you may bring in a single 8.5 x 11 sheet of notes. No calculators or other aides will be permitted. Please bring blue books to the exam. The midterm exam will occur in class on Tuesday, March 10th.

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

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

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 $\sqrt{N}\left(\hat{\theta} - \theta_0\right) \xrightarrow{D} \mathcal{N}\left(0, \Lambda_0\right)$ 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] Calculate a 95 confidence interval for β_0.

[b] 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] We can re-write H_0 as

$$H_0 : C\theta = c$$

Provide the appropriate forms for C and c.

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

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

[2] You've been hired by the Government of Honduras to assess the efficacy of treatment for decompression sickness among lobster divers in La Moskitia. In this region of Honduras lobsters are harvested by divers who, on occasion, get decompression sickness which may result in partial paralysis or worse. You are provided the following table of information about 300 diving accident victims.

		$Y=0$ (No Limp)	$Y=1$ (Limp)
$X=0$ (Untreated)	$W=0$ (Depth $<$ 75')	90	10
	$W=1$ (Depth $\geq$ 75')	10	40
$X=1$ (Treated)	$W=0$ (Depth $<$ 75')	30	20
	$W=1$ (Depth $\geq$ 75')	50	50

[a] What is the probability of a victim walking with a limp conditional on treatment ($X = 1$) and non-treatment ($X = 0$)?

[b] What is the probability of a victim receiving treatment conditional on having dived "deep" ($W = 1$) vs. "shallow" ($W = 0$)?

[c] A government official worries that treatment is harming the divers and thinks it would be better to do nothing. Present a counter-argument to this official.

[d] Let $Y(0)$ and $Y(1)$ denote a divers potential outcome given non-treatment and treatment respectively. Discuss the conditional independence assumption assumption

$$(Y(0), Y(1)) \perp X | W = 0, 1.$$

Make a positive and negative argument for this assumption.

[3] 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:

$$\mu_X = 3, \mu_Y = 2$$

and

$$\sigma_X^2 = 1/2, \sigma_Y^2 = 1/4, \sigma_{XY} = 1/3.$$

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

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

[4] 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] You read in the Oakland Tribune newspaper that Berkeley graduates earn an average of \$55,000 per year nationwide, while the earnings of dropouts average only \$35,000. Express this population earnings difference between Berkeley graduates and dropouts in terms of the statistical model given above.

[b] Under what conditions is it true that $\beta = 20,000$? Do you think these conditions are likely to be true in practice? Briefly explain your answer.

[c] The same article reports that among Berkeley graduates 80 percent come from families where at least one parent completed college, while among all former students (i.e., graduates and dropouts) only 50 percent come from such families. It also states that the overall (i.e., unconditional) graduation rate at Berkeley is 50 percent. What fraction of dropouts come from families where at least one parent completed college?

[d] Assume $\gamma = 10,000$. 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.

[e] 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?

[f] 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?

[5] 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] 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.

[b] Assume that $\theta = -4$. 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] 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 0.9. What is your belief after your null finding?

[6] Consider the following regression model for the log of child height in Brazil (measured as a percentage of the median height of a well-nourished child in the United States of the same age and sex as the sampled child):

$$\begin{aligned}\ln(\text{HEIGHT}) = \alpha_0 + \beta_0 \ln(\textsc{Exp}) + \gamma_0 \text{M}\textsc{Lit} + \delta_0 \text{F}\textsc{Lit} \\ + \zeta_0 \ln(\text{MH}\textsc{eight}) + \eta_0 \ln(\text{FH}\textsc{eight}) + \lambda_0 \text{E}\textsc{nviro} + U\end{aligned}$$

where

$$\mathbb{E}\left[U \middle| \ln\left(\textsc{Exp}\right), \text{ML}\textsc{it}, \text{FL}\textsc{it}, \ln\left(\text{MH}\textsc{eight}\right), \ln\left(\text{FH}\textsc{eight}\right), \text{H}\textsc{ealth}\right] = 0,$$

and $\ln(\textsc{Exp})$ is the log of per capita household expenditure, $\text{ML}\textsc{it}$ equals one if the child's mother is literate and zero otherwise, $\text{FL}\textsc{it}$ equals one if the child's father is literate and zero otherwise, $\ln(\text{MH}\textsc{eight})$ and $\ln(\text{FH}\textsc{eight})$ are the log of mother and father height (also measured as a percentage of the relevant US median) and $\text{E}\textsc{nviro}$ is a measure of how healthy the child's immediate environment is (this might capture things like water quality, sanitation conditions and so on). Thomas and Strauss (1992, *Journal of Development Economics*) fit the shorter regression

$$\ln\left(\text{HEIGHT}\right) = a_0 + b_0 \ln\left(\textsc{Exp}\right) + c_0 \text{ML}\textsc{it} + d_0 \text{FL}\textsc{it} + g_0 \ln\left(\text{MH}\textsc{eight}\right) + h_0 \ln\left(\text{FH}\textsc{eight}\right) + V, \tag{1}$$

where (by construction)

$$\mathbb{E}^*\left[V \middle| \ln\left(\textsc{Exp}\right), \text{ML}\textsc{it}, \text{FL}\textsc{it}, \ln\left(\text{MH}\textsc{eight}\right), \ln\left(\text{FH}\textsc{eight}\right)\right] = 0,$$

to a sample of 36,974 Brazilian children. Their OLS estimates (with standard errors in parentheses) are

$$\begin{aligned}\ln\left(\widehat{\text{HEIGHT}}\right) =\propto \underset{(0.227)}{0.875} \ln\left(\textsc{Exp}\right) + \underset{(0.097)}{0.641} \text{ML}\textsc{it} \\ + \underset{(0.095)}{0.536} \text{FL}\textsc{it} + \underset{(0.846)}{34.43} \ln\left(\text{MH}\textsc{eight}\right) + \underset{(0.888)}{26.11} \ln\left(\text{FH}\textsc{eight}\right)\end{aligned}$$

[a] The Minister of Health in Brazil is interested in assessing the likely effects of providing per capita cash transfers to poor households on child height. Assume that potential recipient households spend their entire monthly income so that such a transfer would augment per capita expenditures one-for-one. Which coefficient, β_0 or b_0, is more relevant to the Minister? Why?

[b] Test the one-sided hypothesis

$$\begin{aligned} H_0 &: b_0 > 0 \\ H_1 &: b_0 \leq 0, \end{aligned}$$

at the five percent level. The 0.95 quantile of a standard normal random variable is approximately 1.645. Show how to calculate the associated p-value.

[c] Thomas and Strauss do not report the covariance in the sampling error of their estimates of g_0 and h_0. Let σ_g^2 be the sampling variance associated with the estimated coefficient on $\ln(\text{MH}\textsc{eight})$, σ_h^2 the sampling variance associated with the estimated coefficient on $\ln(\text{FH}\textsc{eight})$, let σ_{gh} be the (unreported) covariance. The correlation of the two sampling errors is given by

$$\rho = \frac{\sigma_{gh}}{\sigma_g \sigma_h}$$

and is bounded between -1 and 1 by construction. Let $r_0 = g_0 - h_0$ and $\widehat{r} = \widehat{g} - \widehat{h} = 34.43 - 26.11 = 8.32$. Form a 95 percent confidence interval for r_0 as a function of ρ. The 0.975 quantile of a standard normal random variable is approximately 1.96.

[d] Let $\theta = (a, b, c, d, g, h)'$. Express the null $H_0 : r_0 = 0$ as

$$\begin{aligned} H_0 &: C\theta_0 = c \\ H_1 &: C\theta_0 \neq c \end{aligned}$$

for some C and c. Can we reject this null hypothesis for all values of ρ at the 5 percent level?