

Review Sheet 2

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. 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 Thursday, March 23rd.

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

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

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

[a] First show that

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

[b] Second verify the result using the Projection Theorem.

[c] Show that

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

[d] The Vice Chancellor for Undergraduate Education is interested in boosting academic performance among first year students. She random divides first year students into two equal-sized groups. In the first group she randomly assigns half of students to receive a daily snack voucher worth \$5 dollars. In the second group she randomly assigns half of students to get two hours of structured advising each semester. At the end of the semester she records student grade point average. Explain how the Vice Chancellor can use her data to form an estimate of the best linear predictor of end-of-first year GPA given a constant, a dummy variable for snack voucher receipt and a dummy variable for receipt of extra advising.

[e] Under what circumstances is the linear regression computed in part [d] helpful for allocating resources across initiatives? Consider, and elaborate on, three cases: [a] snacks and advising are complements in the production of GPA, [b] they are substitutes and [c] they do not interact.

[f] Outline a more informative experiment for the Vice Chancellor. Explain why is it is “better” than the experiment described in part [d].

[2] The Vice Chancellor for Undergraduate Education is concerned about students dropping out for Cal prior to finishing the requirements for a BA. She provides you with the following Table. The table refers to the Cal students who first arrived on campus in the Fall semester of 2013.

	Number in F13 Still at Cal	Number Dropping out	Number Transferring	Hazard	Survival	Std. Error
F13	6,000	500	200			
S14		530	70			
F14		940	260			
S15		350	150			

The “Number Transferring” column reports the number of students who transfer to another University at the close of the semester. You may assume that these students are lost to further follow-up. The “Std. Error” column refers to the standard error of the survival function.

[a] State and discuss the “random censoring” assumption introduced in lecture. Is this assumption credible in the current context? Explain.

[b] Under the maintained assumption of random censoring fill-in the empty cells in the table. What is the median number of semesters enrolled at Cal prior to drop-out.

[c] The Vice Chancellor provides you with additional information on whether a student is a “first generation” college student. She is concerned that dropout behavior may vary across first generation and non first-generation students. Explain, in detail, how you would conduct a discrete hazard analysis targeted toward this question for the Vice Chancellor.