Prof. James MacKinnon *QUEEN'S UNIVERSITY* *Department of Economics*

Economics 850 Fall, 2024

Assignment 1

Due: September 24, 2024

1. Suppose that X and Y are two binary random variables, each of which can only take on the values 0 and 1. The joint distribution of X and Y is

Table 1. Joint distribution of X and Y

	$Y = 0$	$Y = 1$
$X = 0$	0.32	0.22
$X = 1$	0.16	0.30

a) What is the distribution of Y conditional on $X = 0$? What is the distribution of Y conditional on $X = 1$?

b) What is the expectation of Y conditional on $X = 0$? What is the expectation of Y conditional on $X = 1$?

c) What is the variance of Y conditional on $X = 0$? What is the variance of Y conditional on $X = 1$?

d) What is the unconditional expectation of Y? What is the unconditional variance of Y?

e) Let $Z = XY$. What are the unconditional mean and variance of Z?

[25]

2. Suppose that the $n \times k_1$ matrix $\boldsymbol{X}_1$ contains some of the columns of the $n \times k$ matrix $\boldsymbol{X}$. Let $\boldsymbol{P}_1$ denote the matrix that projects orthogonally onto $\mathcal{S}(\boldsymbol{X}_1)$, and let $\boldsymbol{P_X}$ denote the matrix that projects orthogonally onto $\mathcal{S}(\boldsymbol{X})$.

a) Explain geometrically why $\boldsymbol{P}_1\boldsymbol{P_X} = \boldsymbol{P}_1$.

b) Show that, as a consequence of this result, $\boldsymbol{M}_1\boldsymbol{M_X} = \boldsymbol{M_X}$, where $\boldsymbol{M_X}$ and $\boldsymbol{M}_1$ are the orthogonal complements of $\boldsymbol{P_X}$ and $\boldsymbol{P}_1$, respectively.

c) Show that $\boldsymbol{P_X} - \boldsymbol{P}_1$ is an orthogonal projection matrix. That is, show that this matrix is symmetric and idempotent.

d) Show that the subspace onto which $\boldsymbol{P_X} - \boldsymbol{P}_1$ projects is $\mathcal{S}(\boldsymbol{M}_1\boldsymbol{X}_2)$, where $\boldsymbol{X}_2$ is the $n \times (k - k_1)$ matrix such that $\boldsymbol{X} = [\boldsymbol{X}_1 \;\; \boldsymbol{X}_2]$.

[25]

3. Using a pseudo-random number generator that yields standard normal random numbers, generate 100, 400, 1600, 6400, 25,600, and 102,400 random variates from each of the distributions given below, and calculate the sample means and sample variances of each set of random variates. Also calculate the third central moment around the true mean, if the mean exists. In each case, indicate whether your simulation results suggest that a law of large numbers applies. Are the results, especially for the largest sample size, what you would expect?

a) The normal distribution with mean 1 and variance 16.

b) The Cauchy distribution, recentered and rescaled so that it has median 1 and interquartile range 4 times that of the Cauchy distribution itself; see ETM2, page 38.

c) The Student's t distribution with 3 degrees of freedom, rescaled so that it has mean 1 and variance 16; see ETM2, page 146.

d) The $\chi^2(2)$ distribution, recentered so that it has mean 0.

e) A mixture of two normal distributions, one of them standard normal and one of them normal with mean 2 and variance 16. Choose the mixture probabilities so that the mean of the mixture distribution is 1.

[25]

4. This question uses the file `men2015b.csv`, which is in the data directory on the class website. It contains 32,437 observations on four variables. They are `age` (in years, from 21 to 80), `educ` (education level, which takes five values from 1 to 5), `marry` (binary, equals 1 if the person is currently married and not separated), and `earnings`, weekly earnings in dollars. The data come from the Current Population Survey and are for men in 2015.

a) Create the log of weekly earnings, say `lgearn`. Then regress it on `age`, `age` squared, as many education dummies as possible, and the marriage dummy. How many coefficients can you estimate?

b) Regress `lgearn` on every explanatory variable except `marry`, and regress `marry` on the same set of explanatory variables. Now recover the original coefficient on `marry` by using the FWL theorem. Why is the reported confidence interval slightly shorter?

c) Calculate a measure of leverage and plot it against `age`. What do you learn from this plot? Does every observation for which `age` is the same have the same leverage? Why or why not?

d) Plot leverage against `age` for observations with `marry` = 1 and `educ` = 5. Why does this plot look different from the previous one?

e) Calculate a vector of "omit 1" residuals $\hat{\boldsymbol{u}}^{(\cdot)}$ for the regression. The i^{th} element of $\hat{\boldsymbol{u}}^{(\cdot)}$ is the residual for the i^{th} observation calculated from a regression that uses data for every observation except the i^{th} . Do not actully run this regression 32,437 times! Compare the variance of the omit-1 residuals with the variance of the ordinary ones.