International School of Economics at TSU

Econometrics 2

Lasha Chochua

Problem Set 5

Instructions: You are encouraged to solve the problems before the recitation. Additionally, you are encouraged to work in groups. It is **not mandatory** to submit solutions unless stated otherwise. However, if you would like to share your solution, I would be happy to review it.

Problem 1: Show that:

1. If $e \sim \mathcal{N}(0, I_n\sigma^2)$ and $H'H = I_n$, then $u = H'e \sim \mathcal{N}(0, I_n\sigma^2)$.
2. If $e \sim \mathcal{N}(0, AA')$, then $u = A^{-1}e \sim \mathcal{N}(0, I_n)$.

Problem 2:

Let $X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$ be $N_3(\mu, \Sigma)$ with

$$\mu^T = (2, -3, 1)$$

and

$$\Sigma = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{bmatrix}.$$

(a) Find the distribution of $3X_1 - 2X_2 + X_3$.

Problem 3: Let $\mathbf{X}$ be distributed as $N_3(\mu, \Sigma)$, where

$$\mu^T = (1, -1, 2)$$

and

$$\Sigma = \begin{bmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix}.$$

Which of the following random variables are independent? Explain.

1. X_1 and X_2
2. X_1 and X_3
3. X_2 and X_3
4. X_1 and $X_1 + 3X_2 - 2X_3$

Problem 4: The model is given by:

$$y_i = x_i\beta + e_i, \quad \mathbb{E}(e_i|x_i) = 0$$

where x_i, β, and e_i are scalar. We consider the estimator:

$$\tilde{\beta} = \frac{\bar{y}}{\bar{x}} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i}.$$

We assume that x_i and e_i have finite fourth moments and that $\{y_i, x_i\}$ are a random sample.

a. Find $\mathbb{E}(\tilde{\beta}|X)$.
b. Find $\text{Var}(\tilde{\beta}|X)$.
c. Show that $\tilde{\beta} \to_p \beta$ as $n \to \infty$. Does this require any additional assumptions?
d. Find the asymptotic distribution of $\sqrt{n}(\tilde{\beta} - \beta)$ as $n \to \infty$.

Problem 5: Take the linear model

$$y_i = x_i\beta + e_i$$

$$E(e_i \mid x_i) = 0$$

with n observations and x_i is scalar (real-valued). Consider the estimator

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i^3 y_i}{\sum_{i=1}^n x_i^4}$$

Find the asymptotic distribution of $\sqrt{n}(\hat{\beta} - \beta)$ as $n \to \infty$.

Problem 6: Empirical Exercise Use the provided data **attend.dta** to answer this question.

a. To determine the effects of attending lecture on final exam performance, estimate a model relating *stndfnl* (the standardized final exam score) to *atndrte* (the percent of lectures attended). Include the binary variables *frosh* and *soph* as explanatory variables. Interpret the coefficient on *atndrte*, and discuss its significance.

b. How confident are you that the OLS estimates from part a are estimating the causal effect of attendance? Explain.

c. As proxy variables for student ability, add to the regression *priGPA* (prior cumulative GPA) and *ACT* (achievement test score). Now what is the effect of *atndrte*? Discuss how the effect differs from that in part a.

d. What happens to the significance of the dummy variables in part c as compared with part a? Explain.

e. Add the squares of *priGPA* and *ACT* to the equation. What happens to the coefficient on *atndrte*? Are the quadratics jointly significant?

f. To test for a nonlinear effect of *atndrte*, add its square to the equation from part e. What do you conclude?

Problem 7: Empirical Exercise – OLS Implementation in R

Use the `gpa1` dataset from the `wooldridge` package throughout (accompaning Jupyter Notebook is provided).

- **Part (a).** Load the `wooldridge` package and the `gpa1` dataset. Print all variable names. How many observations does the dataset contain?
- **Part (b).** Construct the dependent variable Y as `colGPA` and the regressor matrix X with a column of ones (intercept), `hsGPA`, and `ACT`. Print the first 6 rows of X and verify its dimensions.
- **Part (c).** Using $\hat{\beta} = (X'X)^{-1}X'Y$, compute the OLS estimates. Report and interpret each coefficient:
  - What is the predicted college GPA for a student with `hsGPA` = 3.0 and `ACT` = 25?
  - Which regressor has a larger effect on college GPA – `hsGPA` or `ACT`? Justify your answer.
- **Part (d).** Compute residuals $\hat{u} = Y - X\hat{\beta}$. Then compute:

$$\hat{\sigma}^2 = \frac{1}{n-k-1}\hat{u}'\hat{u}$$

  Report $\hat{\sigma}$ and interpret its magnitude.

- **Part (e).** Compute the estimated variance-covariance matrix:

$$\widehat{\text{Var}}(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1}$$

  Extract standard errors as square roots of the diagonal elements of $\widehat{\text{Var}}(\hat{\beta})$.

- **Part (f).** Run the same regression using `lm(Y ~ X - 1)` and `summary()`. Compare coefficients and standard errors from Parts (c) and (e) with those reported by `lm()`. Do they match? Why do we use `-1` in the `lm()` formula here?

- **Part (g).** *(Conceptual)* The variance formula in Part (d) divides by $n - k - 1$ rather than n. Using Theorem 2 from the lecture, explain why this choice makes $\hat{\sigma}^2$ an unbiased estimator of σ^2.