Econometrics II: Assignment 2

Due: Thursday, February 20th

1 The r.v. Y is distributed according to the uniform distribution on $[a, b]$,

$$Y \sim \text{unif}[a, b].$$

Assume we observe n i.i.d. realizations $Y_1, \ldots, Y_n$. We are interested in $\mu = \mathbb{E}(Y)$.

For this question, you will need to look up expected value, variance, and covariances of order statistics. I do not expect you to know these by heart, but I leave looking them up as part of the exercise. But to be clear, you may quote these results without proof.

1.1 Show: The OLS estimator is $\hat{\mu}_{\text{OLS}} = \overline{Y}$, the sample average.

(Relatedly, recall that any OLS routine can be [ab]used to computed simple sample averages. How?)

1.2 Consider the alternative estimator that first orders all observations by size, say in increasing order, and then sets

$$\hat{\mu}_{\text{order}} = \frac{Y_1^* + Y_n^*}{2},$$

where $(Y_1^*, \ldots, Y_n^*)$ is the rearranged data set.

Show: (i) This estimator is unbiased. (ii) This estimator has weakly lower variance than $\overline{Y}$, strictly so if $n \geq 3$.

1.3 Does the Gauss-Markov Theorem apply to $\overline{Y}$? If not, why not? If yes, why is it not contradicted?

2 Consider simple linear regression (i.e., X is a scalar) without constant, i.e., the true model is:

$$Y = \beta \cdot X + \varepsilon,$$

though we will also consider estimation of the simple linear regression model

$$Y = \alpha + \beta \cdot X + \varepsilon$$

on the same data. In the following, the slope estimator *without* constant is denoted $\tilde{\beta}$, the estimator with constant is $\hat{\beta}$. We assume i.i.d. sampling and existence of moments etc. as needed.

2.1 Solve for $\tilde{\beta}$ in closed form.

2.2 Characterize and compare mean and variance of $\tilde{\beta}$ and $\hat{\beta}$.

2.3 Clarify that the Gauss-Markov theorem is not contradicted by doing two things:

- Characterize $\mathbb{E}(\tilde{\beta}|\boldsymbol{X})$ *without* assuming that $\alpha \neq 0$.
- State and prove an adapted Gauss-Markov Theorem for the case where we truly know that there is no constant.

3 This question is a reminder that most confidence regions invert hypothesis tests and that the concept of confidence interval is intellectually distinct from the CLT. The fact that most confidence intervals you'll ever see are "estimator ± 2 standard errors" can tend to obscure this.

We want to estimate the efficacy of a vaccine that has been randomly administered to (about) half of a trial sample. Outcomes (i.e., infection of not) were observed a few months later.

3.1 We begin by estimating $\pi \equiv \Pr\{vaccinated \mid infected\}$ from the corresponding sample frequency. Suppose that 178 infections occurred in sample, of these 9 in vaccinated participants. Compute

- the estimator $\hat{\pi}$,
- a 95% Wald (t-statistic based) confidence interval,
- a 95% confidence interval by inverting a Poisson approximation,
- a 95% confidence interval by inverting the binomial distribution.

Note that the last two intervals will be asymmetric. if $F_\pi(x)$ is a hypothesized (approximate) sampling distribution of the sample average as function of true parameter π, you will have to find π^* s.t. $F_{\pi^*}(\overline{x})$ takes on specific values, where $\overline{x}$ is the empirical sample average. Explain.

(You can check your answer to the last part by using an off-the-shelf implementation of Clopper-Pearson testing, e.g. MATLAB's binofit, but please work from scratch first.)

3.2 Prove: If $[\underline{\theta}, \overline{\theta}]$ is a $(1-\alpha)$-confidence interval for θ and $f(\cdot)$ is a known, strictly monotonic function, then $[f(\underline{\theta}), f(\overline{\theta})]$ or $[f(\overline{\theta}), f(\underline{\theta})]$ (depending on direction of monotonicity of $f(\cdot)$) is a $(1-\alpha)$-confidence interval for $f(\theta)$.

3.3 Vaccine efficacy is defined as relative risk reduction, that is,

$$VE \equiv 1 - \frac{\Pr(infected \mid vaccinated)}{\Pr(infected \mid unvaccinated)}.$$

Show: If $\rho \equiv \Pr(vaccinated)$ is known, vaccine efficacy is a strictly monotonic function of π.

3.4 Suppose now that the above numbers came from a trial with 18559 person-years of exposure in the vaccinated group and 18708 years in the control group. (Assignment was originally by flipping fair coins, but this adjusts for attrition and the like. Note that we ignore randomness in this number, which is negligible in this example.) Compute estimators and confidence intervals for VE that correspond to the three methods above.

3.5 The example is not hypothetical. Can you find the numbers we worked with and that you computed in Table 8, page 22, of the document provided with this homework?