ECON3150/4150: Introductory Econometrics – Postponed Exam Spring 2025

Be brief and to the point.
Always motivate your answers.
Use the tables at the end of the exam where necessary.

1. (4 points) True or False? Explain.

   (a) Correlation (between X and Y) does not imply causation (from X to Y).

   ANSWER HINT: *True.* Correlation can arise from reverse causality, omitted variables or pure chance. Additional identification (e.g. random assignment, credible IV) is needed to establish causality.

   (b) Without correlation (between X and Y) there cannot be causation (from X to Y).

   ANSWER HINT: *False.* A causal effect could average out to zero correlation if it is heterogeneous or non-linear (e.g. a treatment that helps half the population and harms the other half equally), or if omitted variable has the same size but the opposite sign of the causal effect.

---

2. [4 points] Table 1 shows summary statistics for a data frame `df` for a categorical variable `x` which takes on the values 1, 2, 3 and a continuous variable `y`. Compute and interpret the point estimates of the following OLS regressions in `R`:

   (a) `feols(y ~ 1, df)`

   ANSWER HINT: Intercept = overall mean of y = **24**.

   (b) `feols(y ~ x1, df)`

   ANSWER HINT: Baseline is groups 2 & 3. Intercept = 27 (mean of those groups). Coefficient on `x1` = $12 - 27 =$ **-15** (difference between group 1 mean and baseline).

   (c) `feols(y ~ x1 + x2, df)`

   ANSWER HINT: Baseline is group 3. Intercept = 30; $\beta_{x1} = 12 - 30 = -18$; $\beta_{x2} = 22 - 30 = -8$.

   (d) `feols(y ~ I(x2 + x3), df)`

   ANSWER HINT: Intercept = group 1 mean = 12; slope = mean of groups 2 & 3 (27) minus 12 = **15**.

Table 1: Table 1. Summary statistics for df

Group (x)	Share	Average y
1	0.2	12
2	0.3	22
3	0.5	30
Total	1.0	24

---

3. [6 points] Consider the following data and regression results which study the relationship between income, education and region:

```
##                              m1             m2
## Dependent Var.:      log(inc)       log(inc)
##
## edu             0.093 (0.001)  0.061 (0.001)
## north                         -0.007 (0.025)
## edu x north                   -0.050 (0.003)
## Constant         9.73 (0.013)   10.2 (0.013)
## _______________ _____________ ______________
## S.E. type                 IID            IID
## R2                      0.406          0.657
## Observations           10,000         10,000
```

(a) State and interpret the point estimate of the coefficient on `edu` in the first regression (`m1`)

ANSWER HINT: $\hat{\beta}_{edu} \approx 0.093$. Interpretation: an additional year of schooling is associated with roughly a **9.3 percent** higher income (because the dependent variable is in logs).

(b) Compute the p-value for the coefficient on `edu` in the first regression. Is it economically and statistically significant?

ANSWER HINT: Standard error 0.001, t-statistic 0.093 / 0.001 93. Two-sided p-value $2\cdot(1-\Phi(93))$. The attached table only goes to $\Phi(2.99)$ which means that the p-value will be much smaller than $2\cdot(1-\Phi(2.99)) = 0.002$, i.e. **highly statistically significant**. A 5 % return per year is also **economically meaningful**.

(c) Suppose you add variable `iq` to the first regression `m1`. Predict qualitatively what will likely happen to the estimate of the coefficient on `edu` and explain why (hint: OVB formula).

ANSWER HINT: IQ is positively correlated with both education and income, so the bias term $\text{Cov}(\text{edu}, \text{iq})\,\beta_{iq}/\text{Var}(\text{edu})$ is positive. Controlling for IQ will **reduce** the education coefficient.

(d) Consider the second regression `m2`. Test the null hypothesis that the coefficient on `edu x north` is zero against the two-sided alternative. What does this tell you about the return to education in the north?

ANSWER HINT: Interaction estimate −0.05 with SE=0.003, t=-16.7, i.e. in absolute value larger than 1.95 which means that we reject H_0 at the 5 percent level. The return to schooling in the North (~1 %) is far lower than elsewhere (~6 %).

(e) What is the expected log-income difference between North and non-North adults who each have 12 years of schooling?

ANSWER HINT: Gap = $\beta_{north} + 12\beta_{edu:north} = -0.607$. Using the regression estimates, this is about −0.61 log-points (exp(-.607)-1 −45 %).

(f) Describe how to obtain a standard error for the difference in part

(e) by running a single additional regression.

ANSWER HINT: Re-centre education at 12:

```
m2c   = feols(log(inc) ~ north*I(edu-12), data = df)
```

The coefficient on `north` in `m2c` **is** the gap at 12 years, and its reported standard error is the required SE.

---

4. [4 points] Revascularization is a medical procedure designed to restore blood flow to tissues or organs when it has been blocked or severely reduced. The following table shows results from Angrist, et al. (2025, NEJM) who analyze a randomized control trial (RCT), that compared conservative and invasive (revascularization) strategies for management of coronary artery disease. Participants were randomly assigned to revascularization, but many assigned to invasive treatment were not revascularized as planned, and participants not assigned revascularization (but to conservative treatment) crossed over to revascularization.

Outcome measure	Control mean	Reduced-form (ITT)	First-stage (compliance)	Per-protocol (as-treated)
SAQ angina-frequency score	90.36	3.69	0.683	3.95
	(15.94)	(0.421)	(0.011)	(0.424)

The outcome here is an angina (chest pain or pressure) frequency score. The first column reports averages and standard deviation (in parentheses). The other columns report reduced form, first-stage estimates, and the per-protocol (as-treated) treatment effect estimates (standard errors in parentheses), the latter which compares participants by treatment received (revascularization vs. conservative treatment).

(a) Interpret the reduced form, first-stage and per-protocol (as-treated) estimates. What do you need to assume for these to be causal?

ANSWER HINT:

- **ITT (3.69):** Assignment to invasive strategy improves the score by 3.69 points. Causal if randomization, no differential attrition, and SUTVA holds.
- **First stage (0.683):** Assignment raises the probability of actually getting revascularization by 0.683. Same assumptions. The instrument is relevant (F is ca 78).
- **As treated (3.95):** Compares those who received the procedure with those who did not, ignoring assignment. Causal only if treatment receipt is exogenous (no unmeasured confounders).

(b) Construct and interpret the instrumental variable estimate of the effect of revascularization. What do you need to assume for this to be causal?

ANSWER HINT: IV = 3.69 / 0.683 **5.40**. Interpretation: Average effect of people who switch treatment status because of the assignment (compliers), receiving revascularization increases the angina-frequency score by 5.4 points. Causal if in addition to the assumptions above the **exclusion restriction** restriction holds: assignment only affects the outcome through the treatment. Seems uncontroversial here.