cannot estimate model → β is not unique

need to drop

Multicollinearity : some x in linear function of another x

1st identify term in linear function

2nd drop some of them

$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + U$

$X_1 = aX_2$ → same r.v. but different scale

$X_1 = kX_3$ → drop X_1 or X_3

$X_1 = a + bX_2$ → drop constant, X_1, or X_2 (β_0, $\beta_1 X_1$, $\beta_2 X_2$)

$X_1 = a + bX_2 + cX_3$

care only linear function

$X_2 = X_1^2$

but it's multicollinear if X_1 if dummy

dummy: $0 \to 0^2 = 0$, $1 \to 1^2 = 1$

Independence → $P(X|Y) = P(X)$

can use this to say that if there's no mean inde, then there's no indepen

unbiased also imply consistent

$P(x_i)$ only OLS estimator

$E[X|Y] = \sum_i x_i P(x_i|Y) = \sum_i x_i P(x_i) = E[X]$

$E[U|X] = 0$ ⇒ $cov(U,X) = 0$

mean independence → $E[U|X] = 0$ → $cov(U,X) = 0$

uncorrelated

$p \wedge q$ $p \leftrightarrow q$ $p \vee q$

inde, mean, not mean, not independent

$p \to q \equiv \neg q \to \neg p$

T	T	≡ T
T	F	≡ F
F	T	≡ T
F	F	≡ T

⇒ if there's no mean inde, then there's no independent as well

---

Midterm: 10 pages, write inside box only → can use pencil, want very short, concise answer not essay

40 scores part 1 : written test

40 part 2&3 : use 2B

- part 2 : 10 questions true or fail, if you don't have enough info → false
- part 3 : 15 questions 4 choices → let you answer twice

* will tell clearly the interpretation < correlation / causation

OLS = justify a little bit or add control

Exam design:

EASY 20% 30%

MEDIUM : require some calculation → Hypo test 40% - 60%

not new CI

HARD : don't see before 20 - 30%

CI but use 2-sided test

└ given 1-side-test → can you do 2-side test

mean ~40-50% last yr

Lecture 1

- probability / distribution
  - marginal, joint, conditional
- moments, expectation - calculate from distribution
  - property E[ ] to simplify
  - Var[ ] = $E[X^2] - (E[X])^2$
  - SD( )
  - Cov( )
  - var(ax+by)
  - $var[x+y] = var[x] + var[y] + 2cov(x,y)$
- understand independent, mean inde, uncorrelate
- good (random) estimator
  - prove:
    - unbiased
    - consistent

Lecture 2 linear regression model: ☆ $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$

1st model construction !!!

- interprete β → impact of what?
- ln, level

2nd estimation

able to read result from STATA ⇒ reg, gen, test (understand r.v.)

3rd inference

- hypothesis test
- confidence interval

Rescale

Goodness of fit: no need to remember formula