**Практическая работа «Инструментальные переменные»**

**Часть 1: Экономическое развитие и институты**

**Часть 1. Обсуждение модели**

Ваша задача — исследовать взаимосвязь между экономическим развитием и институтами.

**Исследовательский вопрос**

1. Сформулируйте исследовательский вопрос.
2. Сформулируйте исследовательскую гипотезу.
3. Обсудите экономический механизм, который бы ее объяснял.

**Модель**

1. Обсудите модель парной линейной регрессии, которая подходит для описания изучаемых взаимосвязей.
2. Обсудите проблему эндогенности. К каким последствиям она может привести?
3. Обсудите, какие инструментальные переменные необходимо включить в модель для решения проблемы эндогенности.

**Набор данных**

На этом семинаре мы работаем с материалами статьи **Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. American economic review, 91(5), 1369-1401**. Статья доступна по [ссылке](https://www.aeaweb.org/articles?id=10.1257/aer.91.5.1369).

Используемые в статье данные и код доступны [онлайн](https://www.openicpsr.org/openicpsr/project/112564/version/V1/view). Данные загружены в папку Teams в формате txt.

Загрузите и изучите данные. Обсудите, какие переменные лучше всего подходят для ответа на поставленный исследовательский вопрос.

Описание переменных

| **Name** | **Label** |
| --- | --- |
| longname | full country name |
| shortnam | 3 letter country name |
| logmort0 | Original Log Settler Mortality |
| risk | Exp Risk |
| loggdp | Log GDP per capita (PPP) in 1995 |
| campaign | Original mortality from campaign |
| source0 | Country source of own m. rate |
| slave | Data from forced labor |
| latitude | Absolute value of latitude |
| neoeuro | ‘New-Europes’(USA,CAN,AUS,NZL) |
| asia | Asia |
| africa | Africa |
| other | ‘Other’ Cont (AUS, NZL, MLT) |
| edes1975 | Pop of Euro Descent, 1975 |
| campaignsj | Ошибка:509 |
| campaignsj2 | Ошибка:509 |
| mortnaval1 | mortality using naval stations, method 1 |
| logmortnaval1 | logr mortality using naval stations, method 1 |
| mortnaval2 | mortality using naval stations, method 2 |
| logmortnaval2 | log mortality using naval stations, method 2 |
| mortjam | mortality benchmarking to the Caribbean |
| logmortjam | log mortality benchmarking to the Caribbean |
| logmortcap250 | log mortality capped at 250 |
| logmortjam250 | log mortality benchmarked to the Caribbean, capped at 250 |
| wandcafrica | west and central Africa |
| malfal94 | Falciparum malaria index 1994 |
| wacacontested | contested observations in west and central Africa |
| mortnaval2250 | mortality using naval stations, capped at 250 |
| logmortnaval2250 | log mortality using naval stations capped at 250 |
| mortnaval1250 | mortality using naval stations, capped at 250 |
| logmortnaval1250 | log mortality using naval stations capped at 250 |

**Разведывательный анализ данных**

Проведите разведывательный анализ данных:

1. Постройте таблицу с описательными статистиками основных переменных.
2. Постройте графики, визуализирующие изучаемую взаимосвязь.

**Регрессионный анализ**

Проведите регрессионный анализ:

1. Оцените базовую модель парной регрессии. Рассчитайте робастные стандартные ошибки.
2. Оцените изучаемую взаимосвязь с использованием инструментальных переменных и двухшагового метода наименьших квадратов:

* используйте функцию ivreg() из пакета AER, рассчитайте робастные стандартные ошибки.
* попробуйте повторить результаты, используя функцию lm().

1. Сведите результаты оценки всех регрессий в одну таблицу.
2. Проинтерпретируйте полученные результаты.
3. Оформите скрипт R так, чтобы он был понятным и воспроизводимым.

**Бонусы**

1. Пробуйте ориентироваться на материалы статьи и “воспроизвести” графики и таблицы, которые используют авторы.
2. Обсудите, как можно было бы оценить подобные модели для случая множественной регрессии. Произведите оценивание в R.

**Часть 2: Заработная плата и образование**

Ваша задача — исследовать взаимосвязь между заработной платой и образованием.

**Исследовательский вопрос**

1. Сформулируйте исследовательский вопрос.
2. Сформулируйте исследовательскую гипотезу.
3. Обсудите экономический механизм, который бы ее объяснял.

**Модель**

1. Обсудите ММЛР, которая подходит для описания изучаемых взаимосвязей.
2. Обсудите проблему эндогенности. К каким последствиям она может привести?
3. Обсудите, какие инструментальные переменные необходимо включить в модель для решения проблемы эндогенности.

**Набор данных**

На этом семинаре мы работаем с материалами исследования **Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling**. Работа доступна по [ссылке](https://www.nber.org/papers/w4483).

Данные для воспроизведения результатов исследования доступны в папке Teams в формате .txt. Это выборка из обследования молодых американцев (National Longitudinal Survey of Young Men). Обследование проводилось дважды: в 1966 и 1976 гг.

Также данные и другие материалы доступны на [странице автора](http://davidcard.berkeley.edu/data_sets.html).

Загрузите и изучите данные. Обсудите, какие переменные лучше всего подходят для ответа на поставленный исследовательский вопрос.

| **Name** | **Label** |
| --- | --- |
| id | sequential id runs from 1 to 5225 |
| nearc2 | grew up near 2-yr college |
| nearc4 | grew up near 4-yr college |
| nearc4a | grew up near 4-yr public college |
| nearc4b | grew up near 4-yr priv college |
| ed76 | education in 1976 |
| ed66 | education in 1966 |
| age76 | age in 1976 |
| daded | dads education missing=avg |
| nodaded | 1 if daded imputed |
| momed | moms education |
| nomomed | 1 if momed imputed |
| weight | nls weight for 1976 cross-section |
| momdad14 | 1 if live with mom and dad age 14 |
| sinmom14 | lived with single mom age 14 |
| step14 | lived step parent age 14 |
| reg661 | dummy for region=1 in 1966 |
| reg662 | dummy for region=2 in 1966 |
| reg663 | dummy for region=3 in 1966 |
| reg664 | dummy for region=4 in 1966 |
| reg665 | dummy for region=5 in 1966 |
| reg666 | dummy for region=6 in 1966 |
| reg667 | dummy for region=7 in 1966 |
| reg668 | dummy for region=8 in 1966 |
| reg669 | dummy for region=9 in 1966 |
| south66 | lived in south in 1966 |
| work76 | worked in 1976 |
| work78 | worked in 1978 |
| lwage76 | log wage (outliers trimmed) 1976 |
| lwage78 | log wage in 1978 outliers trimmed |
| famed | mom-dad education class 1-9 |
| black | 1 if black |
| smsa76r | in smsa in 1976 |
| smsa78r | in smsa in 1978 |
| reg76r | in south in 1976 |
| reg78r | in south in 1978 |
| reg80r | in south in 1980 |
| smsa66r | in smsa in 1966 |
| wage76 | raw wage cents per hour 1976 |
| wage78 | raw wage cents per hour 1978 |
| wage80 | raw wage cents per hour 1980 |
| noint78 | if noninterview in 1978 |
| noint80 | if noninterview in 1980 |
| enroll76 | 1 if enrolled in 1976 |
| enroll78 | 1 if enrolled in 1978 |
| enroll80 | 1 if enrolled in 1980 |
| kww | the kww score |
| iq | a normed iq score |
| marsta76 | mar status in 1976 1=married, spouse present |
| marsta78 | mar status in 1978 1=married, spouse present |
| marsta80 | mar status in 1980 1=married, spouse present |
| libcrd14 | 1 if library card in home age 14 |

**Разведывательный анализ данных**

Проведите разведывательный анализ данных:

1. Постройте таблицу с описательными статистиками основных переменных.
2. Постройте графики, визуализирующие изучаемую взаимосвязь.
3. Сделайте выводы о качестве данных.
4. Обсудите, можно ли провести разведанализ для подтверждения силы инструментов? Годности? Если да, то как его провести?

**Регрессионный анализ**

Проведите регрессионный анализ:

1. Оцените базовую ММЛР (рекомендация на основе статьи: log(wagei)=β0+β1educi+β2experi+β3exper2i+β4blacki+β5smsai+ui

Проинтерпретируйте результаты.

1. Добавьте в модель образование матери.

* Что показывает коэффициент при новой переменной? Имеет ли он ожидаемый знак? Значим ли этот коэффициент в ваших расчетах?
* Чему равна отдача от образования в данном случае? Как она изменилась по сравнению с ответом в предыдущем пункте?

1. Добавьте в модель кросс-эффект образования матери и образования самого респондента.

* Что показывает коэффициент при новой переменной? Имеет ли он ожидаемый знак? Значим ли этот коэффициент в ваших расчетах?
* Чему равна отдача от образования в данном случае? Как она изменилась по сравнению с ответом в предыдущем пункте?
* Проверьте гипотезу о значимости эффекта образования в зависимости от образования матери.

1. Оцените изучаемую взаимосвязь с использованием инструментальных переменных и двухшагового метода наименьших квадратов:

* Используйте функцию ivreg() из пакета AER.
* Как она изменилась по сравнению с ответами в предыдущих пунктах?

1. Сведите результаты оценки всех регрессий в одну таблицу.
2. Проинтерпретируйте полученные результаты.
3. Оформите скрипт R так, чтобы он был понятным и воспроизводимым.