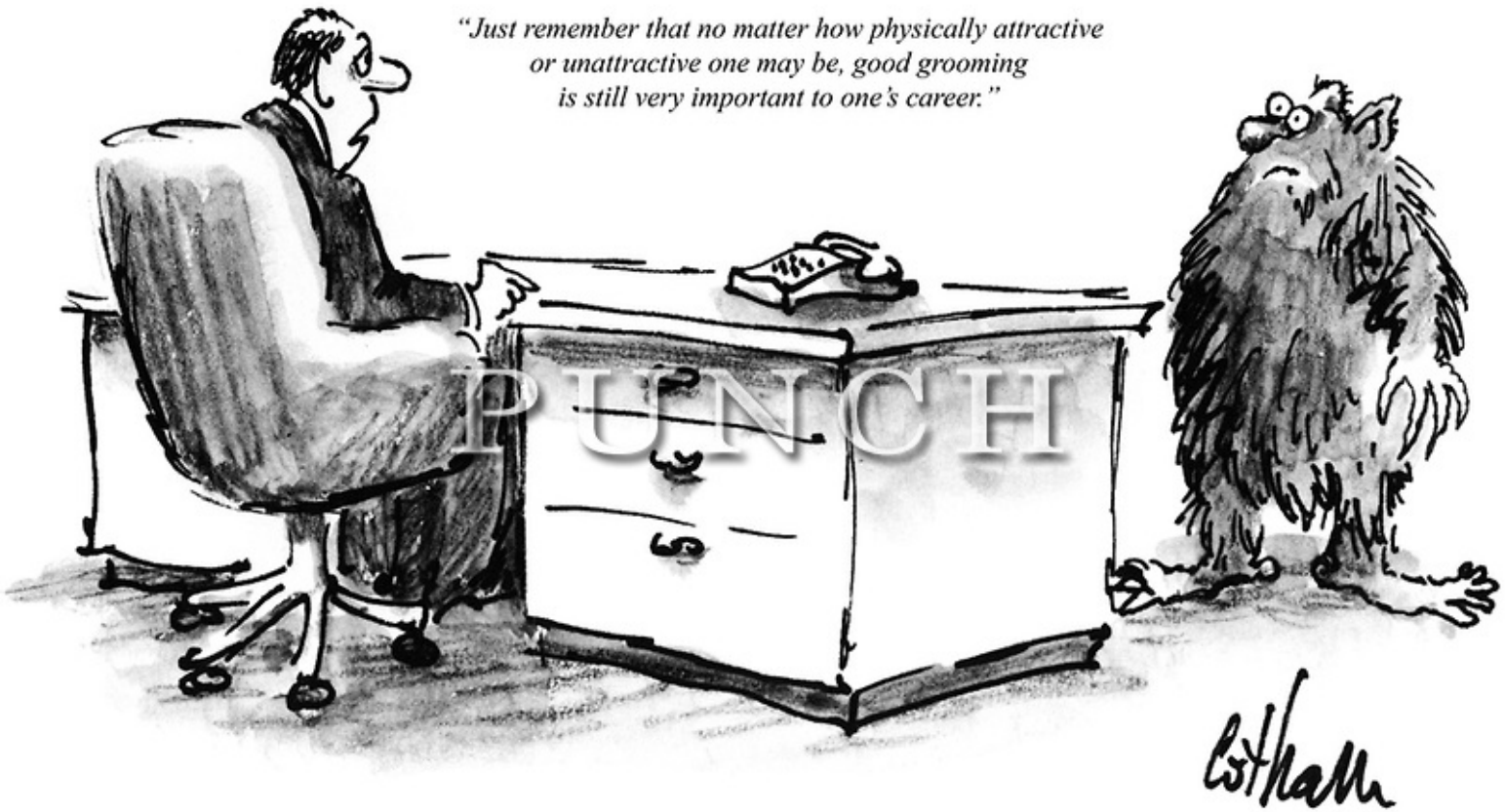
Do attractive people get paid more?



Econometrics: Computer Modelling

Felix Pretis

Programme for Economic Modelling
Oxford Martin School, University of Oxford

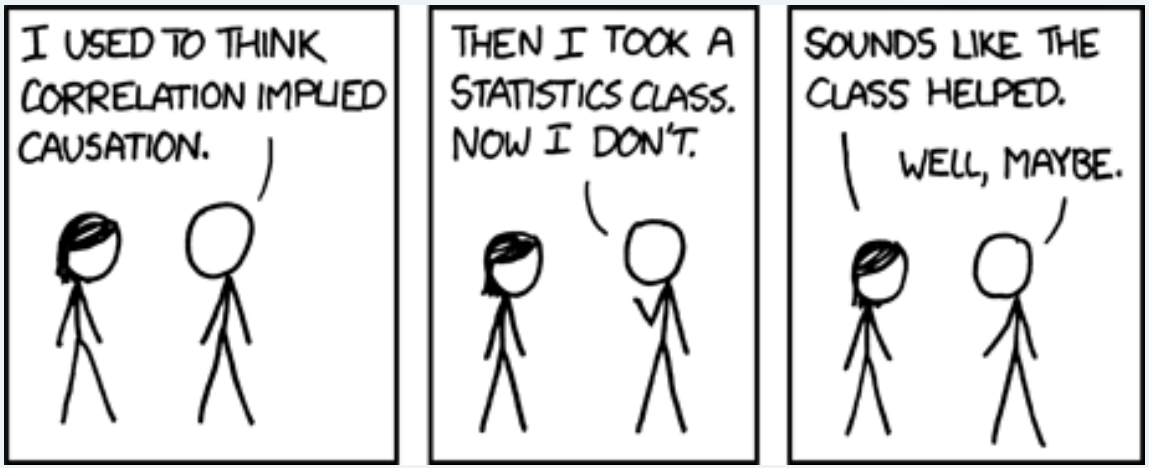
Lecture 1: Introduction to Econometric Software & Cross-Section Analysis

Aim:

- Introduce econometric modelling in practice
- Introduce OxMetrics/PcGive Software

By the end of the course:

- Able to build econometric models
- Evaluate output and test theories
- Use OxMetrics/PcGive to load, graph, model, data



- **Textbooks:** no single text book. Useful:
  - Doornik, J.A. and Hendry, D.F. (2013). *Empirical Econometric Modelling Using PcGive 14: Volume I*, London: Timberlake Consultants Press.
    - Included in OxMetrics installation – "Help"
  - Hendry, D. F. (2015) *Introductory Macro-econometrics: A New Approach*.
    - Freely available online: `http://www.timberlake.co.uk/macroeconometrics.html`
- **Lecture Notes & Lab Material online:** `http://www.felixpretis.org`
- **Problem Set**: to be covered in tutorial
- **Exam**: Questions possible (Q4 and Q8 from past papers 2016 and 2017)

- 1: **Intro to Econometric Software & Cross-Section Regression**
- 2: Micro-Econometrics: Limited Indep. Variable
- 3: Macro-Econometrics: Time Series

- Economies high dimensional, interdependent, heterogeneous, and evolving: comprehensive specification of all events is impossible.
- Economic Theory
  - likely wrong and incomplete
  - **meaningless** without empirical support
  - Econometrics to discover new relationships from data
  - Econometrics can provide empirical support. . . or refutation.
- Require econometric software unless you really like doing matrix manipulation by hand.

Structure of data

	wage	hours	IQ	KWW	educ	exper	tenure	age	married
1	769	40	93	35	12	11	2	31	1
2	808	50	119	41	18	11	16	37	1
3	825	40	108	46	14	11	9	33	1
4	650	40	96	32	12	13	7	32	1
5	562	40	74	27	11	14	5	34	1
6	1400	40	116	43	16	14	2	35	1
7	600	40	91	24	10	13	0	30	0
8	1081	40	114	50	18	8	14	38	1
9	1154	45	111	37	15	13	1	36	1
10	1000	40	95	44	12	16	16	36	1
11	930	43	132	44	18	8	13	38	1
12	921	38	102	45	14	9	11	33	1
13	900	45	125	40	15	4	3	30	0
14	1318	38	119	24	16	7	2	28	1
15	1792	40	118	47	16	9	9	34	1
16	958	50	105	37	10	17	2	35	1
17	1360	45	109	39	15	6	9	36	1
18	850	40	72	36	11	19	10	38	1
19	830	44	105	29	14	4	7	29	1
20	471	44	101	34	12	13	7	31	1
21	1275	40	123	37	14	9	1	31	1
22	1615	50	113	49	16	10	4	36	1

Many options!

- **Menu-driven**
  - OxMetrics
  - STATA
  - EViews
  - ...
- **Simple batch programming language**
  - OxMetrics: batch-file
  - STATA: do-file
  - EViews: programme files
  - ...
- **Matrix-based programming languages**
  - R, Ox, Matlab, Python, C, Gauss,...

- Computer program for working with economic data.
  - Front-end: Menu driven, easy to use.
  - But: Not restricted to menus: All underwritten by **Ox Programming language**.
- OxMetrics acts as an umbrella for many **packages**:
  - All have different names: PcGive, G@RCH, PcNaive, ...
  - Different packages for different types of economic data.

OxMetrics Online Help

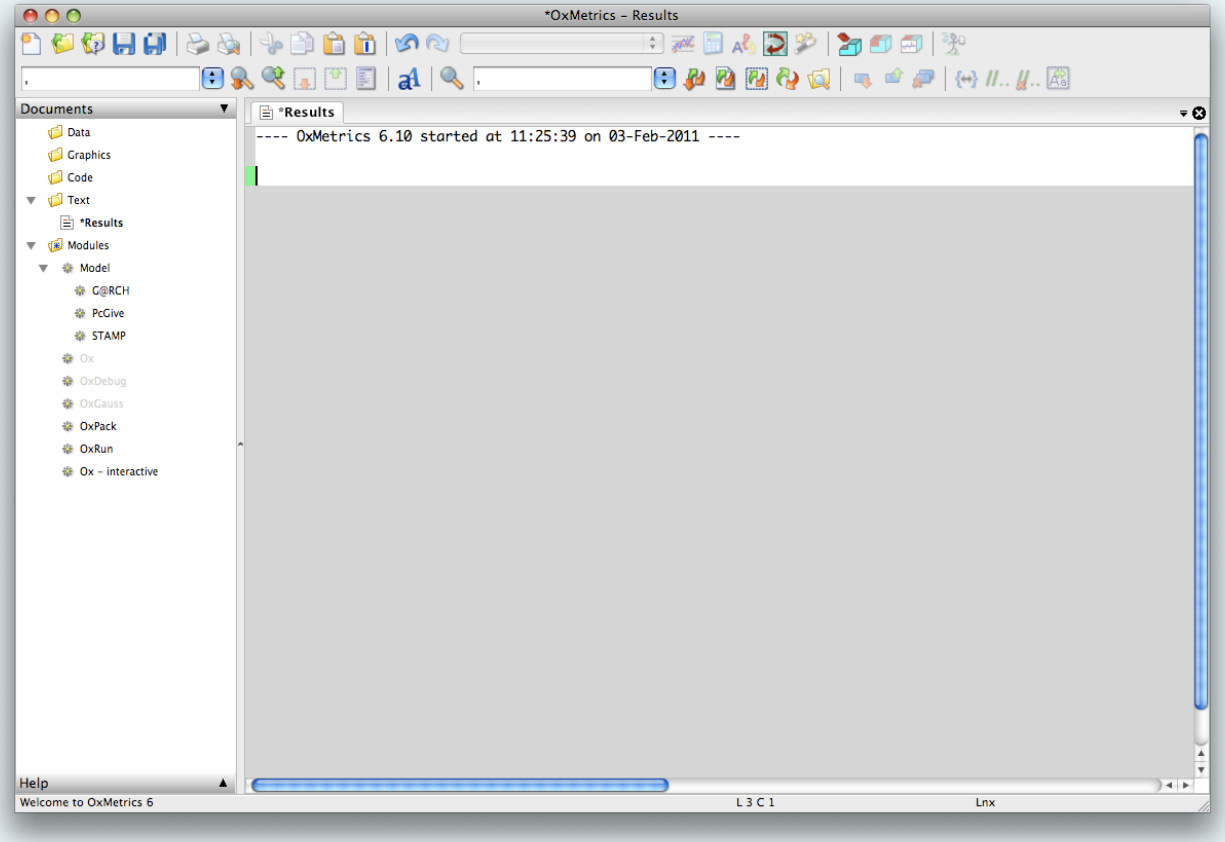
- Online OxMetrics help:
  `http://www.pcgive.com/oxmetrics.html`
- Online PcGive help:
  `http://www.pcgive.com/pcgive/index.html`

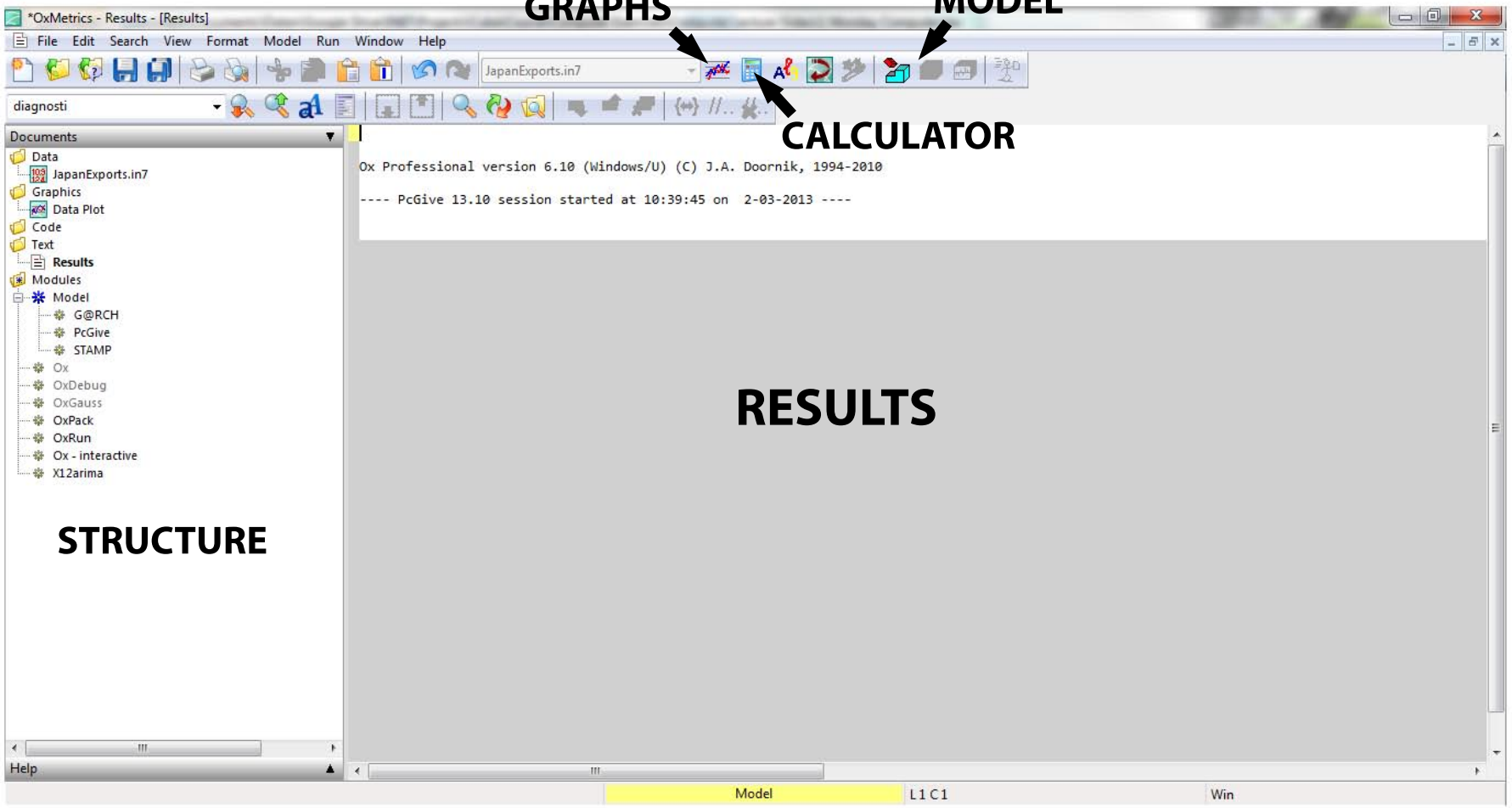
Download OxMetrics onto your own computers from:

`http://www.doornik.com/download/Oxford/`

- Free license for one year
- Download from within Oxford network

- Open OxMetrics on your system. 





→ **Layout of Window**: Structure, Results, Menus

What determines wages?

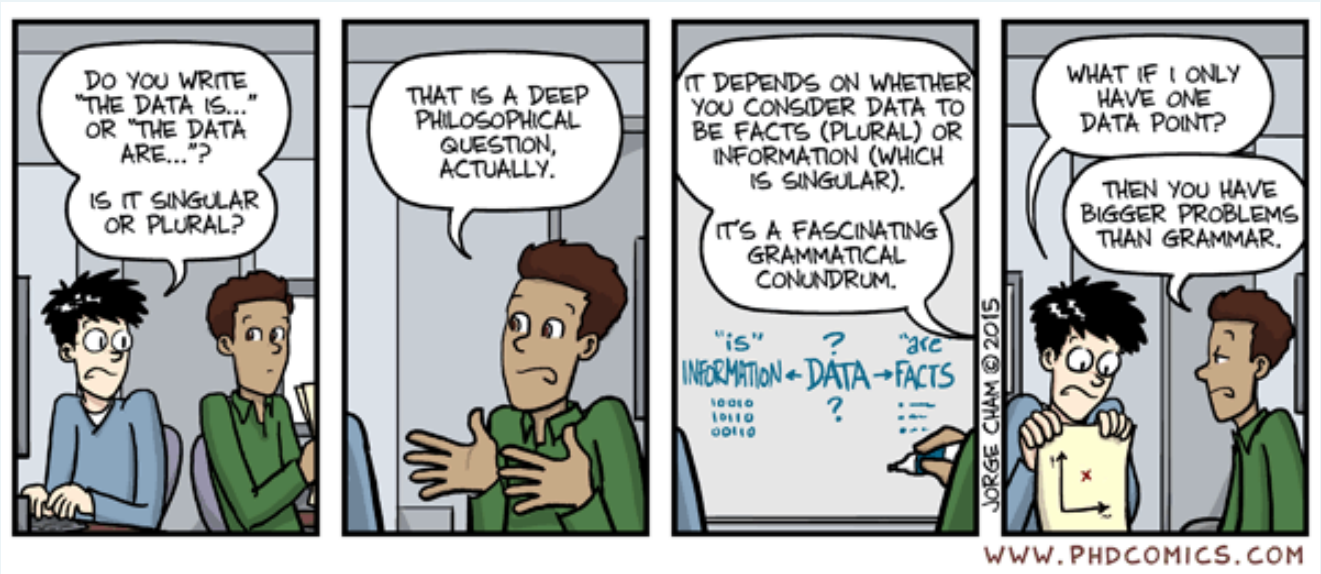
Education?
Ability
...Attractiveness?

Download the Data

- `wages.in7`
- Subset from: Blackburn & Neumark (1992). “Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials”. *Quarterly Journal of Economics*. Vol. 107 (`goo.gl/Ysi9r6`)

Load the Wages Data

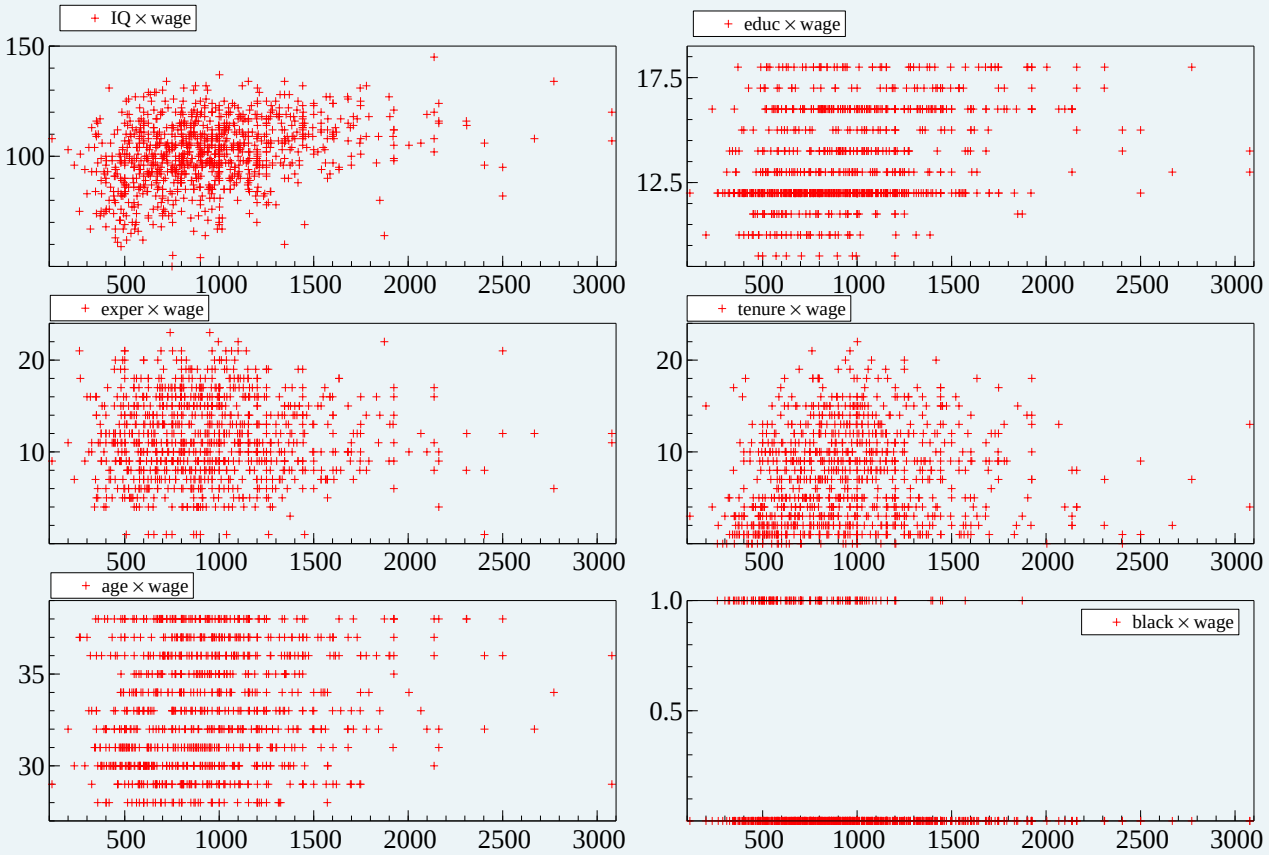
- Open File



- OxMetrics has great flexibility for plotting data.
- Huge range of possible types of plot. Easy to access.
- Multiple series on a set of axes, multiple sets of axes.
- Copy and paste works in wonderful ways. . .
- Can save in pdf format, ideal for including in LaTeX or Word documents.

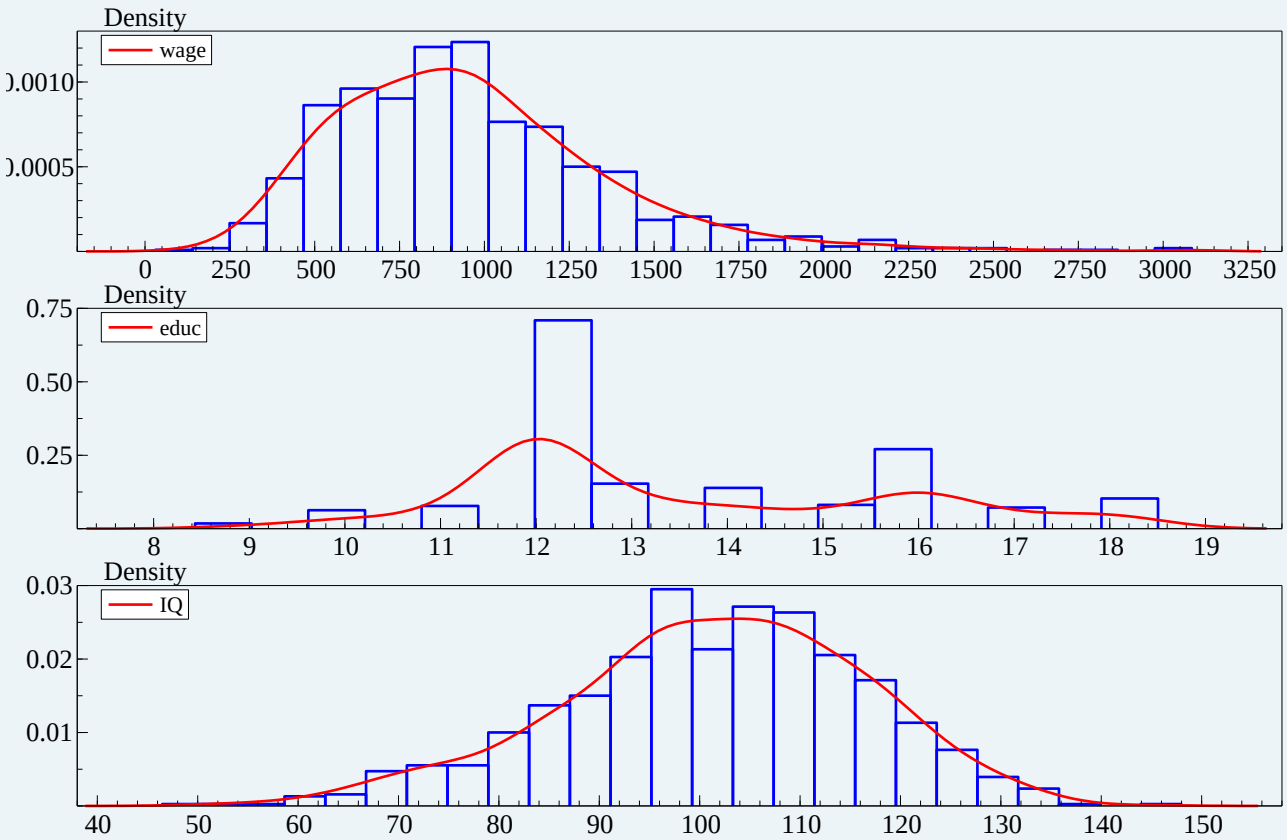
Always plot your Data

- as *Scatter Plots*, *Histograms*, etc.
- Modify the graphs!



Always plot your Data

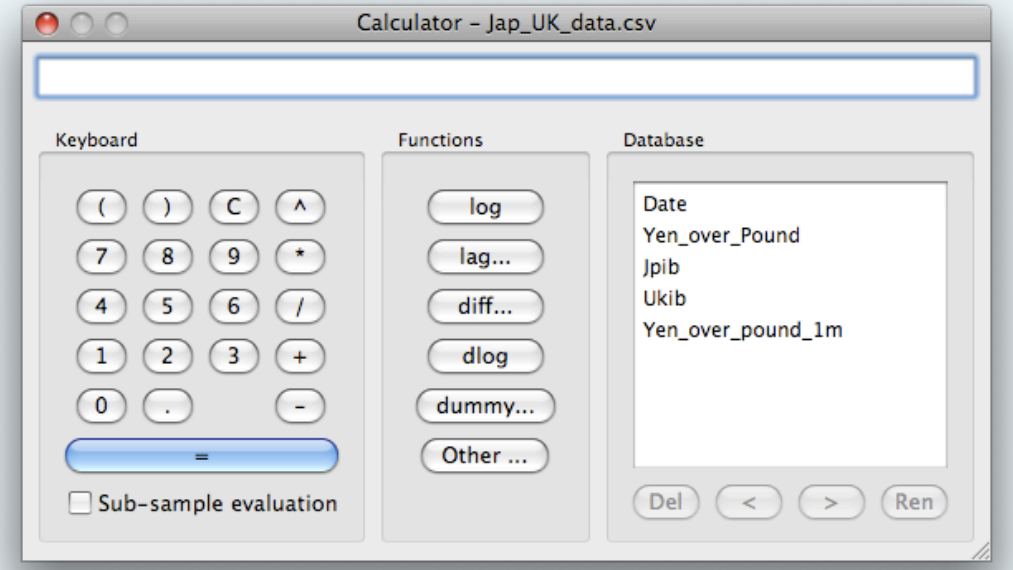
- as *Scatter Plots*, *Histograms*, etc.
- Modify the graphs!



Creating New Variables: Calculator

- Economic theory often proposes objects of interest which are combinations:
  - E.g. Average Propensity to Save: Savings/Income.
  - E.g. Real exchange rate: (Spot rate $\times$ Domestic prices)/Foreign prices.
  - Total Sales = Sales_1 + Sales_2...
  - log(Wages)
- Hence we need to be able to create such variables in OxMetrics!
  - We use the Calculator tool:
    - `Model` and `Calculator`.
    - `Alt+C`.

Calculator Tool



- Data transformations: Lags, logs, differences, percentage changes,...
- `Other...`: Extensive list of data transformations.

→ **Do: Construct log(wages) variable**

Understanding the Data & Building an Econometric Model:

1. Other: Descriptive Statistics (Summary Statistics etc.)
   - Task: Create summary statistics for ”Wages” dataset

2. Cross Section Regression (Regression Model)
   - Build simple model of log(wages): regress log(wages) on education and an intercept

- Create summary statistics:

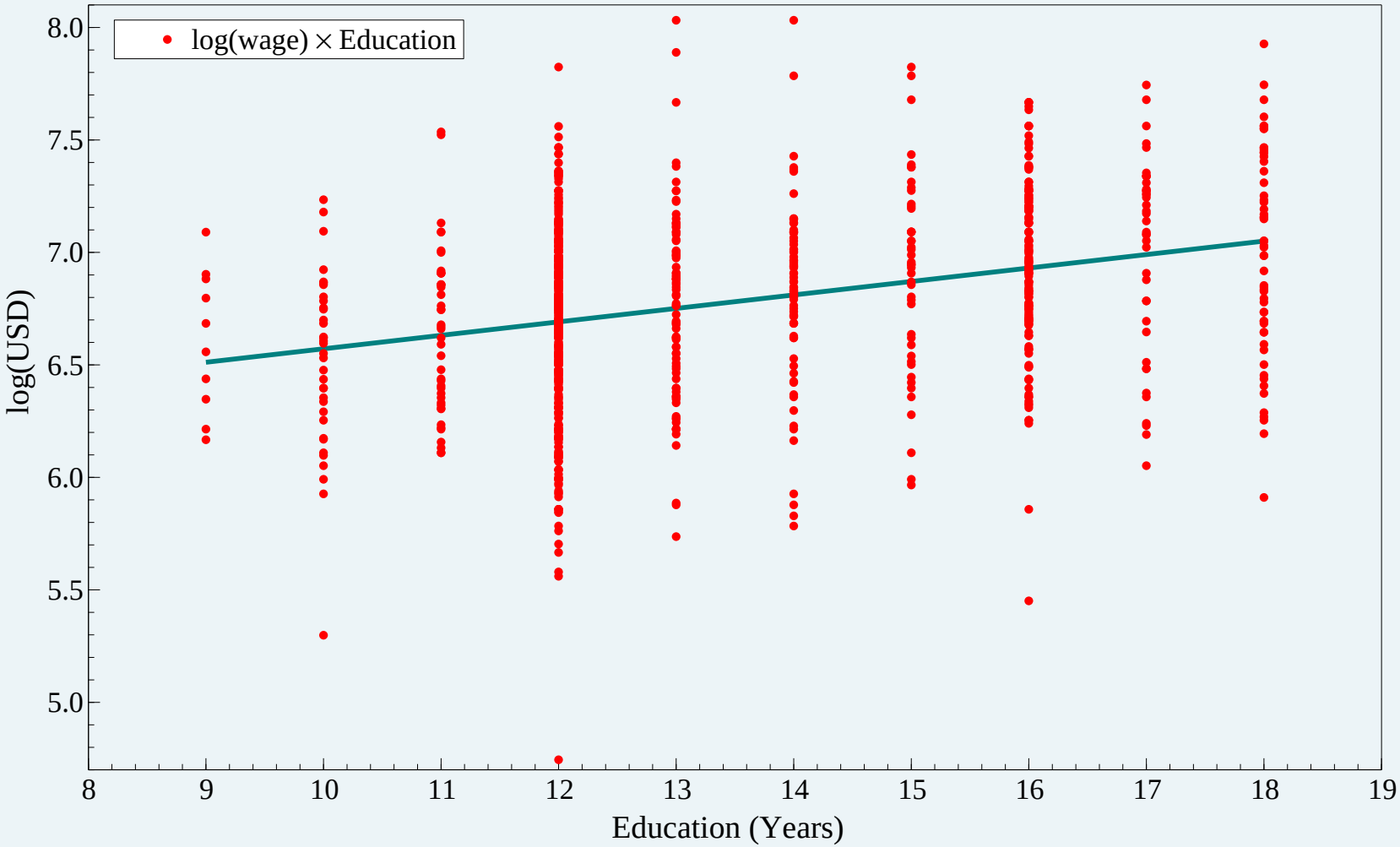
```
Means, standard deviations and correlations
The sample is: 1 - 935 (935 observations and 6 variables)
Means
         wage        hours           IQ          KWW         educ        exper
       957.95       43.929       101.28       35.744       13.468       11.564
Standard deviations (using T-1)
         wage        hours           IQ          KWW         educ        exper
       404.36       7.2243       15.053       7.6388       2.1967       4.3746
Correlation matrix:
               wage        hours           IQ          KWW         educ        exper
wage         1.0000   -0.0095043      0.30909      0.32613      0.32711    0.0021897
hours    -0.0095043       1.0000     0.073839      0.11389     0.091009    -0.062126
IQ           0.30909     0.073839       1.0000      0.41352      0.51570     -0.22491
KWW          0.32613      0.11389      0.41352       1.0000      0.38813     0.017452
educ         0.32711     0.091009      0.51570      0.38813       1.0000     -0.45557
exper     0.0021897    -0.062126     -0.22491     0.017452     -0.45557       1.0000
```

Interpreting Regression Output

```
EQ( 1) Modelling Lwage by OLS-CS
       The dataset is: wages.in7
       The estimation sample is: 1 - 935

                    Coefficient  Std.Error  t-value  t-prob Part.R^2
Constant                5.97306    0.08137     73.4  0.0000   0.8524
educ                  0.0598392   0.005963     10.0  0.0000   0.0974

sigma                   0.40032  RSS                 149.518582
R^2                   0.0974168  F(1,933) =    100.7 [0.000]**
Adj.R^2               0.0964494  log-likelihood       -469.721
no. of observations          935  no. of parameters           2
mean(Lwage)                6.779  se(Lwage)            0.421144
```


log(wage) × Education
log(USD)
Education (Years)
8.0
7.5
7.0
6.5
6.0
5.5
5.0
8
9
10
11
12
13
14
15
16
17
18
19

Omitted variable-bias:

- **Our model**: $\text{L(wage)}_i = \beta_0 + \beta_1 \text{Educ}_i + \nu_i$
- What if: **Data Generating Process "Truth"**= $\text{L(wage)}_i = \beta_0 + \beta_1 \text{Educ}_i + \beta_2 \text{Ability}_i + \epsilon_i$

Then $\hat{\beta}_1$ biased unless $\beta_1 = 0$ or Educ. & Ability are uncorrelated.

Demonstration: **Monte Carlo Simulation**. Suppose

- $\beta_1 = 1$ (Education)
- $\beta_2 = 4$ (Ability, IQ)

With:

- A: Corr(Education, Ability)=0.5
- B: Corr(Education, Ability)=0

Concerns about omitted variables

Start **general**, then reduce to **specific**.

Determinants of Wages – build a more general regression model, including:

- IQ (proxy for ability)
- Experience
- Age
- Race (black)

More general model


```
       The estimation sample is: 1 - 935

                  Coefficient  Std.Error  t-value  t-prob Part.R^2
Constant              5.04194     0.1719     29.3  0.0000   0.4809
educ                0.0517412   0.007549     6.85  0.0000   0.0481
IQ                 0.00454855   0.001037     4.39  0.0000   0.0203
exper               0.0132632   0.003844     3.45  0.0006   0.0127
age                 0.0134906   0.004827     2.79  0.0053   0.0083
black               -0.156964    0.04056    -3.87  0.0001   0.0159

sigma                0.381616  RSS                135.290875
R^2                  0.183304  F(5,929) =     41.7 [0.000]**
Adj.R^2              0.178908  log-likelihood       -422.974
no. of observations       935  no. of parameters           6
mean(Lwage)             6.779  se(Lwage)            0.421144
```

Testing Hypotheses on Model Parameters:

- Exclusion restrictions: *Test - Exclusion Restrictions*
- General restrictions: *Test - General Restrictions*

Do:

- Excluding Age & Black?
- Coefficient on Exper = Coefficient on Age?
- Coefficient on Educ = 0.05?

Testing for Heteroskedasticity:

- White Test for Heteroskedasticity (without cross-products)

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_1^2 + \delta_4 x_2^2 \cdots + \text{error} \qquad (1)$$

  Test: $\delta_1 = 0, \delta_2 = 0, \ldots$

Need to:

- Store residuals from regression $\hat{u}$
- Generate new variables: $\hat{u}^2, x_1^2, \ldots$
- Estimate White model & test exclusion

Need to:

- Store residuals from regression $\hat{u}$ – *Test - store Residuals...*
- Generate new variables: $\hat{u}^2, x_1^2, \ldots$ – *Calculator Function*
- Estimate White test model & test exclusion – as before

```
Test for excluding:
[0] = educ
[1] = IQ
[2] = exper
[3] = age
[4] = black
[5] = educ_sq
[6] = IQ_sq
[7] = exper_sq
[8] = age_sq
Subset F(9,925)   =   0.55046 [0.8378]
```

Luckily OxMetrics reports many diagnostic tests automatically:

	Coefficient	Std.Error	t-value	t-prob	Part.R^2
Constant	5.04194	0.1719	29.3	0.0000	0.4809
educ	0.0517412	0.007549	6.85	0.0000	0.0481
IQ	0.00454855	0.001037	4.39	0.0000	0.0203
exper	0.0132632	0.003844	3.45	0.0006	0.0127
age	0.0134906	0.004827	2.79	0.0053	0.0083
black	-0.156964	0.04056	-3.87	0.0001	0.0159

sigma	0.381616	RSS	135.290875
R^2	0.183304	F(5,929) =	41.7 [0.000]**
Adj.R^2	0.178908	log-likelihood	-422.974
no. of observations	935	no. of parameters	6
mean(Lwage)	6.779	se(Lwage)	0.421144

Normality test:	Chi^2(2) =	28.678 [0.0000]**
Hetero test:	F(9,925) =	0.55046 [0.8378]
Hetero-X test:	F(15,919) =	0.59513 [0.8803]
RESET23 test:	F(2,927) =	0.58041 [0.5599]

Compare: Subset F(9,925) = 0.55046 [0.8378]

Luckily OxMetrics reports many diagnostic tests automatically:

	Coefficient	Std.Error	t-value	t-prob	Part.R^2
Constant	5.04194	0.1719	29.3	0.0000	0.4809
educ	0.0517412	0.007549	6.85	0.0000	0.0481
IQ	0.00454855	0.001037	4.39	0.0000	0.0203
exper	0.0132632	0.003844	3.45	0.0006	0.0127
age	0.0134906	0.004827	2.79	0.0053	0.0083
black	-0.156964	0.04056	-3.87	0.0001	0.0159

sigma	0.381616	RSS	135.290875
R^2	0.183304	F(5,929) =	41.7 [0.000]**
Adj.R^2	0.178908	log-likelihood	-422.974
no. of observations	935	no. of parameters	6
mean(Lwage)	6.779	se(Lwage)	0.421144

Normality test:	Chi^2(2)	=	28.678 [0.0000]**
Hetero test:	F(9,925)	=	0.55046 [0.8378]
Hetero-X test:	F(15,919)	=	0.59513 [0.8803]
RESET23 test:	F(2,927)	=	0.58041 [0.5599]

What about other diagnostic tests?

Education and experience matters, but what about **attractiveness**?

OxMetrics & PcGive Exercise: Attractiveness and Wages

- Load Data: ”attractiveness.in7”
  - “looks”: variable from 1-5 scored looks
  - “attr_belavg”: dummy variable below average attractiveness
  - “attr_abvavg”: dummy variable above average attractiveness
- Graph Data
- Build model of log(Wages)

Hamermesh & Biddle (1994). “Beauty and the Labour Market” *American Economic Review*. Vol. 84 (5) (`goo.gl/3bP3IA`)

Answering the following questions:

1. What measure of attractiveness is most sensible to include?
2. What proportion of the population are rated as above/below average?
3. Is there a significant difference in above/below average attractiveness between men and women? (Hint: think about a regression model that would allow you to test this!)
4. Does attractiveness have an effect on wages? Quantify any effect you find.
5. Is the effect different for men/women?
6. Does the effect of attractiveness change when controlling for other variables? What about other interaction effects?
7. Is your model well-specified?
8. Is there evidence for Heteroskedasticity in your models?
9. What about other diagnostic criteria (Normality etc.)?