

計量経済学入門

麻生良文

講義計画

1. データ分析入門（2回程度）
 2. 確率・統計の基礎（1回程度）
 3. 単回帰（3回程度）
 4. 重回帰（2回程度）
 5. 重回帰　ダミー変数，仮説検定，分散不均一性（3回程度）
 6. Qualitative Response Model（1回程度）
 7. 操作変数法（1回程度）
 8. パネル・データの分析（2回程度）
- 講義で使用するデータセットはWooldridgeの教科書で使われているものを使います。

参考文献

- 入門的な統計学の教科書
 - 東京大学教養学部統計学教室編『統計学入門』東京大学出版会
 - 大屋幸輔『コア・テキスト 統計学』新世社
- 計量経済学（初級）
 - 田中隆一『計量経済学の第一歩』有斐閣
- 計量経済学（中級）
 - Jeffrey M. Wooldridge, Introductory Econometrics, Thomson South-Western
 - J.H.Stock and M.W.Watson, Introduction to Econometrics, Pearson
 - 西山慶彦・新谷元嗣・川口大司・奥井亮『計量経済学』有斐閣
- 計量経済学（上級）
 - W.Greene, Econometric Analysis, Prentice Hall
 - Jeffrey M. Wooldridge, Econometric Analysis of Cross Section and Panel Data, MIT Press
 - Bruce E. Hansen, Econometrics, Princeton Univ. Press

参考文献(2) R関連

- G.Grolemund 『RstudioではじめるRプログラミング入門』 オライリー・ジャパン, オーム社, 2015年, 2600円＋税
- A.ジュール,E.イノウエ,E.ミースターズ 『R初心者のためのABC』 シュプリンガー・ジャパン, 2010年, 3500円＋税
- Norman Matloff 『アート・オブ・Rプログラミング』 オライリー・ジャパン, 2012年, 3200円＋税
- C.Kleiber and A.Zeileis, “Applied Econometrics with R”, Springer, 2008
- Florian Heiss, *Using R for Introductory Econometrics*, 2016
<http://www.urfie.net/> (左のページを参照 Amazo.co.jpでも購入できる)
- Rの入門的文献はネット上に多数あり
<https://cran.r-project.org/index.html> を参照

データ分析入門

- 計量経済分析の方法
- 理論モデル（明確なモデルが無い場合もある）

$$y = f(x_1, x_2, \dots, x_k)$$

- 統計的検証

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u$$

一般的には上のような線型モデルを仮定し，観察されたデータから方程式のパラメータを推計（非線型モデルの推定もある）

u 誤差項(error term), 攪乱項(disturbance term)

y 被説明変数(explained variable), 従属変数(dependent variable)

x 説明変数(explanatory variable), 独立変数(independent variable)

- β に関して線型の方程式の当てはめ（実はそれほど制限的ではない）

データ分析入門(2)

- 統計的検証
 - 推定された係数 b (前頁の β の推計値) の値は理論モデルの予測と矛盾しないか
 - 誤差項の確率分布の仮定 $\rightarrow b$ の信頼区間
- どのように当てはめるか
 - 最小二乗法(least square method)
 - 最尤法(maximum likelihood method)
 - その他(モーメント法など)
 - $\rightarrow$ この講義ではまず最小二乗法をとりあげる
- 入門的な段階では、線型モデル、単一方程式モデルを扱うが次のようなモデルを推計する場合もある
 - 非線形モデル
 - 連立方程式モデル

データ分析入門(3) データの構造

- クロスセクション・データ

例) 家計調査 (ある1時点における様々な世帯 (個人) の所得・消費のデータ)

- 時系列データ

例) GDPの推移、株価の推移

- パネル・データ

例) 同一の個人 (or 企業) を複数時点にわたり追跡した調査
クロスセクション×時系列

都道府県や市町村の所得・人口等のデータを複数時点にわたり追跡した調査→「疑似パネルデータ」とよばれる (集計データであり、完全に個人ベースのデータではないので)

- この講義では、主にクロスセクションデータの分析手法を取りあげる。また、最後の方でパネルデータを取りあげる

計量分析の手順

- 理論モデル，仮説の提示
- データの収集
 - Excelファイルに保存しておくのが便利
- データを統計ソフトにimport
- データの要約
 - 記述統計，グラフ，散布図，ヒストグラム
 - データのおおまかな特徴をとらえる；異常値、入力ミスなどの発見のためにも必要
- 回帰分析
- モデルの診断
 - 残差のチェック，その他（回帰分析の前提チェック）
- 仮説検定

次のページ以降は実際の分析手順のイメージです

excelでまとめたデータセット

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	wage	educ	exper	tenure	nonwhite	female	married	numdep	smsa	northcen	south	west	construc	ndurn
2	3.1	11	2	0	0	1	0	2	1	0	0	1	0	
3	3.24	12	22	2	0	1	1	3	1	0	0	1	0	
4	3	11	2	0	0	0	0	2	0	0	0	1	0	
5	6	8	44	28	0	0	1	0	1	0	0	1	0	
6	5.3	12	7	2	0	0	1	1	0	0	0	1	0	
7	8.75	16	9	8	0	0	1	0	1	0	0	1	0	
8	11.25	18	15	7	0	0	0	0	1	0	0	1	0	
9	5	12	5	3	0	1	0	0	1	0	0	1	0	
10	3.6	12	26	4	0	1	0	2	1	0	0	1	0	
11	18.18	17	22	21	0	0	1	0	1	0	0	1	0	
12	6.25	16	8	2	0	1	0	0	1	0	0	1	0	
13	8.13	13	3	0	0	1	0	0	1	0	0	1	0	
14	8.77	12	15	0	0	0	1	2	1	0	0	1	0	
15	5.5	12	18	3	0	0	0	0	1	0	0	1	0	

Rstudio の画面

Excel等で整理したデータを統計ソフトにimportする

The screenshot displays the RStudio environment with the following components:

- Environment Pane:** Shows the Global Environment with three data objects: `wage1` (526 obs. of 24 variables), `wage1.1m` (List of 12), and `wage2.1m` (List of 12).
- Files Pane:** Displays the file structure for the `linearHypothesis` package, including `linearHypothesis (car)` and `R Documentation`.
- Documentation Pane:** Shows the `Test Linear Hypothesis` documentation, including a description of the `linearHypothesis` function and its usage.
- Console:** Shows the R startup message and the workspace loaded from `~/RData`.

The main data table (`wage1`) is displayed below:

	wage	educ	exper	tenure	nonwhite	female	married	numdep
1	3.10	11	2	0	0	1	0	2
2	3.24	12	22	2	0	1	1	3
3	3.00	11	2	0	0	0	0	2
4	6.00	8	44	28	0	0	1	0
5	5.30	12	7	2	0	0	1	1
6	8.75	16	9	8	0	0	1	0
7	11.25	18	15	7	0	0	0	0
8	5.00	12	5	3	0	1	0	0
9	3.60	12	26	4	0	1	0	2
10	18.18	17	22	21	0	0	1	0
11	6.25	16	8	2	0	1	0	0
12	8.13	13	3	0	0	1	0	0

要約統計量

データのおおまかな特徴をとらえる

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins Project: (None)

Console Terminal Jobs

~/
R is free software and comes with absolutely no warranty.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[workspace loaded from ~/.RData]

> summary(wage1)
      wage      educ      exper      tenure      nonwhite      female      married
Min.   : 0.530   Min.   : 0.00   Min.   : 1.00   Min.   : 0.000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
1st Qu.: 3.330   1st Qu.:12.00   1st Qu.: 5.00   1st Qu.: 0.000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
Median : 4.650   Median :12.00   Median :13.50   Median : 2.000   Median :0.0000   Median :0.0000   Median :1.0000
Mean   : 5.896   Mean   :12.56   Mean   :17.02   Mean   : 5.105   Mean   :0.1027   Mean   :0.4791   Mean   :0.6084
3rd Qu.: 6.880   3rd Qu.:14.00   3rd Qu.:26.00   3rd Qu.: 7.000   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
Max.   :24.980   Max.   :18.00   Max.   :51.00   Max.   :44.000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000

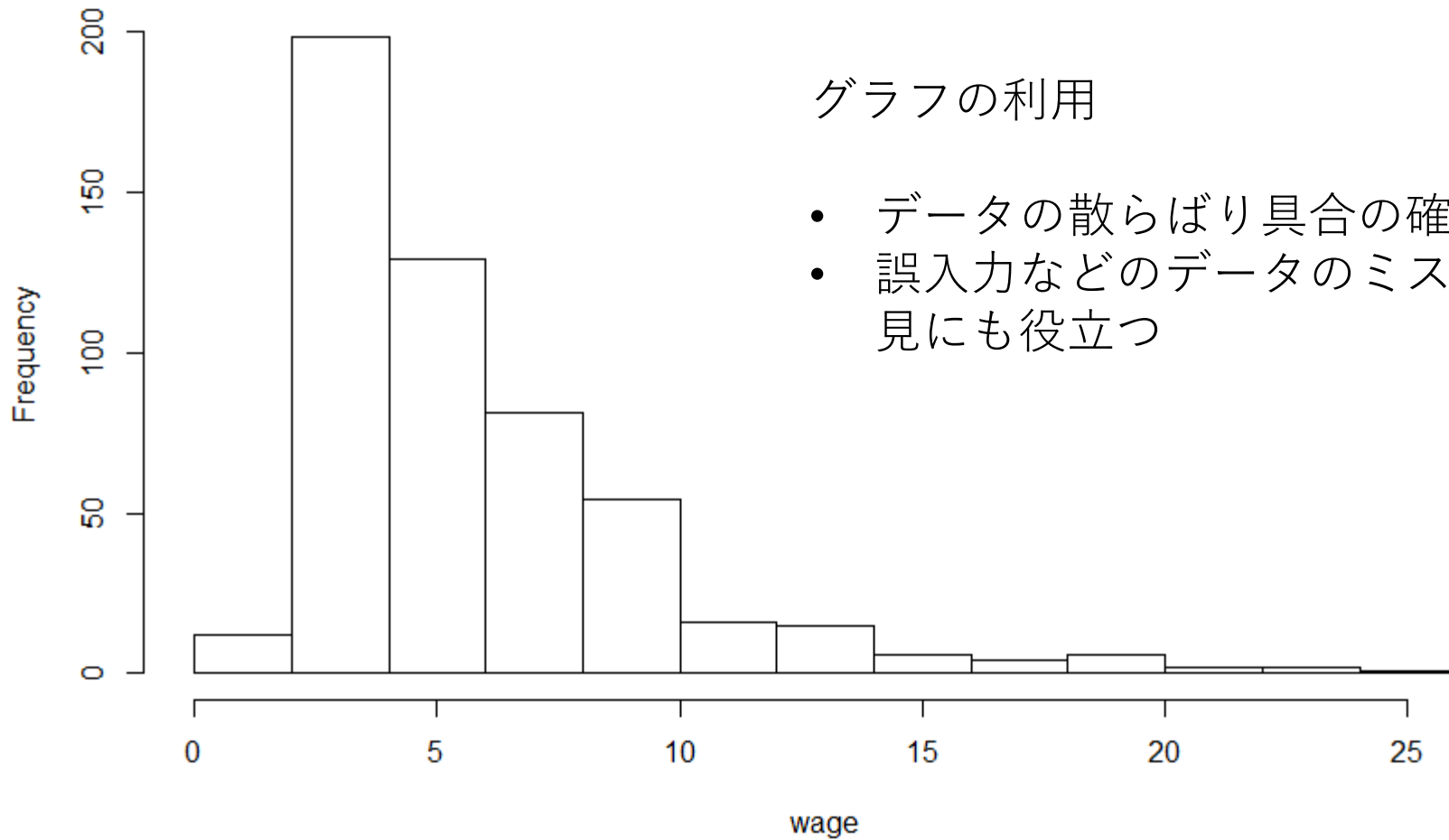
      numdep      smsa      northcen      south      west      construc      ndurman
Min.   :0.000   Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000
1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000
Median :1.000   Median :1.0000   Median :0.000   Median :0.0000   Median :0.0000   Median :0.00000   Median :0.0000
Mean   :1.044   Mean   :0.7224   Mean   :0.251   Mean   :0.3555   Mean   :0.1692   Mean   :0.04563   Mean   :0.1141
3rd Qu.:2.000   3rd Qu.:1.0000   3rd Qu.:0.750   3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.0000
Max.   :6.000   Max.   :1.0000   Max.   :1.000   Max.   :1.0000   Max.   :1.0000   Max.   :1.00000   Max.   :1.0000

      trcommpu      trade      services      profserv      profocc      clerocc      servocc
Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
Median :0.00000   Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
Mean   :0.04373   Mean   :0.2871   Mean   :0.1008   Mean   :0.2586   Mean   :0.3669   Mean   :0.1673   Mean   :0.1407
3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:0.0000
Max.   :1.00000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000

      lwage      expersq      tenursq
Min.   :-0.6349   Min.   : 1.0   Min.   : 0.00
1st Qu.: 1.2030   1st Qu.: 25.0   1st Qu.: 0.00
Median : 1.5369   Median :182.5   Median : 4.00
Mean   : 1.6233   Mean   :473.4   Mean   : 78.15
3rd Qu.: 1.9286   3rd Qu.:676.0   3rd Qu.:49.00
Max.   : 3.2181   Max. :2601.0   Max. :1936.00

> |
```

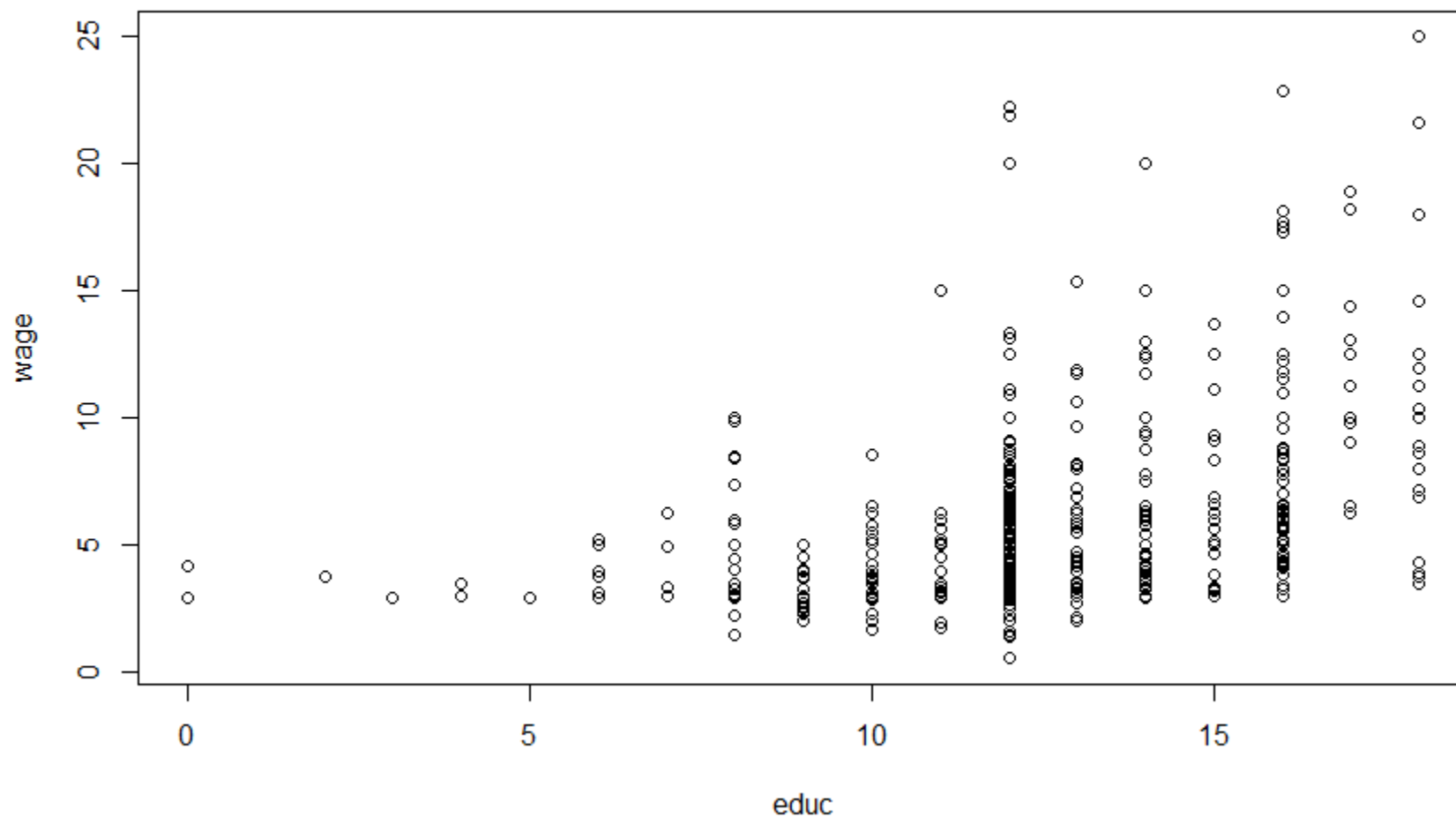
Histogram of wage



グラフの利用

- データの散らばり具合の確認
- 誤入力などのデータのミスの発見にも役立つ

教育年数 (educ) と賃金(wage)の散布図



ここでは教育が賃金に与える効果が分析の目的

回帰分析の結果（Rの出力） 単回帰

Call:
lm(formula = lwage ~ educ)

被説明変数: lwage 賃金の対数値
説明変数: educ 教育年数

Residuals:

Min	1Q	Median	3Q	Max
-2.21158	-0.36393	-0.07263	0.29712	1.52339

回帰分析や出力結果の読み方については、次回以降の講義で解説

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.583773	0.097336	5.998	3.74e-09 ***
educ	0.082744	0.007567	10.935	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4801 on 524 degrees of freedom

Multiple R-squared: 0.1858, Adjusted R-squared: 0.1843

F-statistic: 119.6 on 1 and 524 DF, p-value: < 2.2e-16

回帰分析の結果（Rの出力） 重回帰

Call:
lm(formula = lwage ~ educ + exper + tenure)

被説明変数: lwage(賃金の対数値)
説明変数: educ(教育年数), exper(仕事の経験年数), tenure(勤続年数)

Residuals:

Min	1Q	Median	3Q	Max
-2.05802	-0.29645	-0.03264	0.28788	1.42809

重回帰モデルとは複数の要因の効果を分離して推計する統計的手法

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.284359	0.104190	2.729	0.00656 **
educ	0.092029	0.007330	12.555	< 2e-16 ***
exper	0.004121	0.001723	2.391	0.01714 *
tenure	0.022067	0.003094	7.133	3.29e-12 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4409 on 522 degrees of freedom

Multiple R-squared: 0.316, Adjusted R-squared: 0.3121

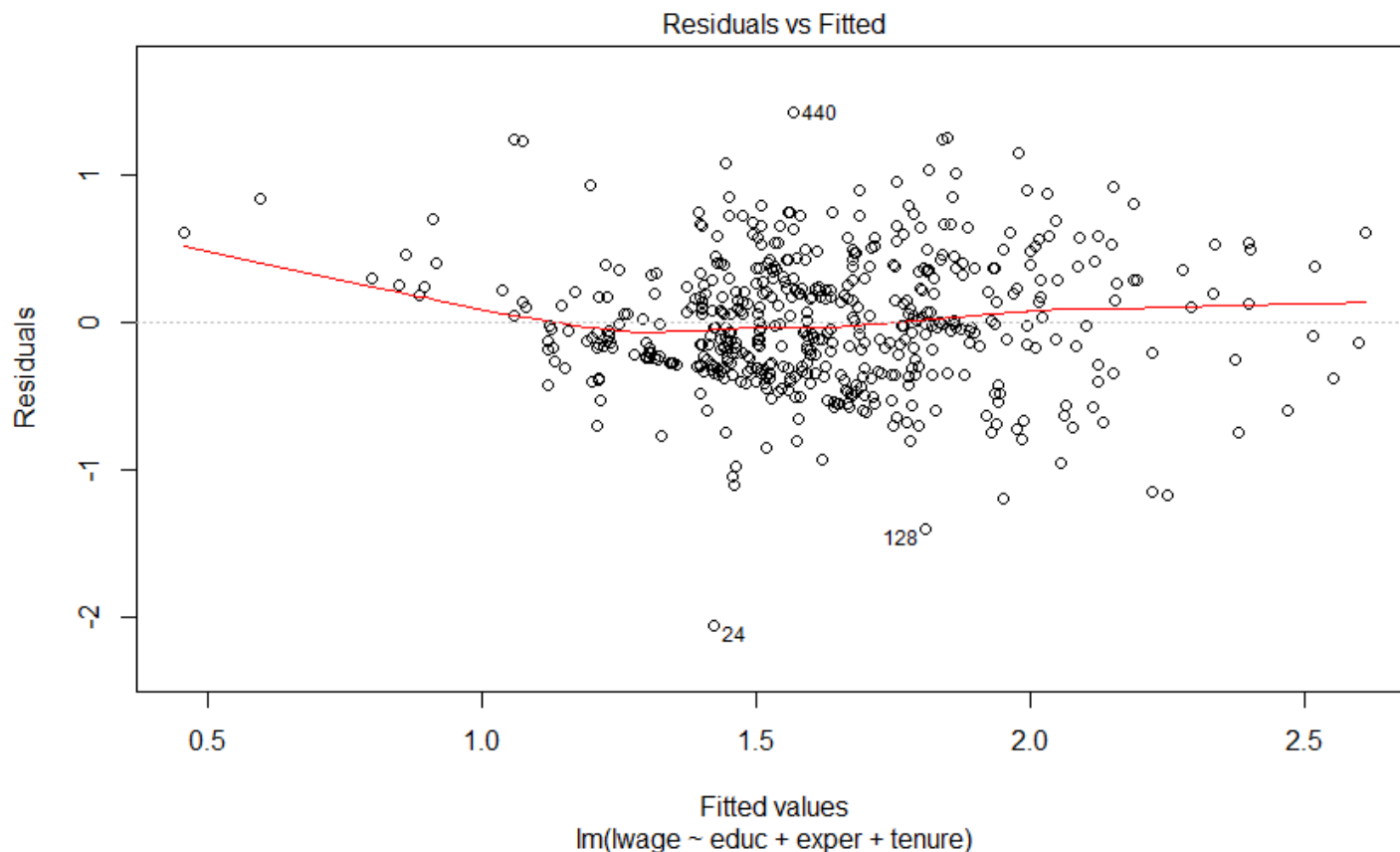
F-statistic: 80.39 on 3 and 522 DF, p-value: < 2.2e-16

残差の検討

回帰分析では、誤差項 u の性質にある前提を置く

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$$

この前提が満たされないと、統計的検証の正しさが保証されない→残差をみることで、誤差項に関する前提が満たされているかどうかをチェックする



データ分析の実際

- データの入力, インポート
- 変数の作成
 - 対数変換, 説明変数の平方, ダミー変数の作成
- 記述統計
 - 異常値のチェック
 - データの大まかな性質をとらえておく
 - みせかけの相関
- 回帰分析
 - 当てはまり具合, 残差のチェック, 定式化の再検討
 - 仮説検定
 - 結果の解釈
 - 理論モデルの再検討

実際のデータ分析では、データの誤入力がよくあります。データベース等から電子データをダウンロードした場合でも、元のデータに誤りがある場合があります。

ヒストグラム、散布図等のグラフ作成や、要約統計量等の出力は、こうしたデータのエラーを見つける上でも重要です。