**STATA DATA ANALYSIS FOR**

**ASSIGNMENTDATA FILE**

**INTRODUCTION**

# Linear Regression Analysis

Programming Language: Python, R

Linear regression is a linear approach to model the relationship between a scalar response (dependent varaible) and one or more explanatory variables (independent variables). The case of having single explanatory variable, the method is referred as simple linear regression. In case of having multiple explanatory variablea, the method is referred as multiple linear regression. Ordinary least squares (OLS) is a type of linear least squares method for estimating the unknown parameters in a linear regression model. OLS chooses the parameters of a linear function of a set of explanatory variables by using the principle of least squares that minimizes the sum of the squares of the residuals" (differences between the observed dependent variable and those predicted by the linear function). The method is largely applied in Econometrics, Finance, Data Science and other subject areas.

**LITERATURE REVIEW**

# Event Study Analysis

Programming Language: STATA, Python

Event studies examine the behavior of firms’ stock prices around corporate events. Event studies focusing on announcement effects for a short-horizon around an event provide evidence relevant for understanding corporate policy decisions.Event studies also serve an important purpose in capital market research as a way of testing market efficiency. In this STATA do file you can find the code of an entire event study analysis that investigates the impact of external event on (e.g. Trump Tweets) on the stock prices, where for the stock prices "daily" data frequency has been considered and for the event window intervals of [-100, 100] days & [-250, 250] days have been considered. The analysis contains stock prices for 45 unique company's. Per company we have run OLS regression where company's return has been regressed on the market return and event date. The do file consists of the following parts:

- Running OLS regressions for 45 companies

- Predicting residuals of the OLS model

- Using obtained residuals to build Empirical Model (GARCH(1/1) & ARCH(1/1))

- Obtaining test statistics to test the hypothesis with KS-test (Kolmogorov-Smirnov test)

Publications:

- Coutts, A. J., Terence, C., Roberts, J., (1994). "The market model and the event study method: A synthesis of the econometric criticisms" International Review of Financial Analysis,3(2): 149-171

# Fixed Effects Regression for Panel Data

Panel data allows you to control for variables you cannot observe or measure like cultural factors or difference in business practices across companies; or variables that change over time but not across entities (i.e. national policies, federal regulations, international agreements, etc.). This is, it accounts for individual heterogeneity. One of the techniques used to analyze Panel Data is Fixed Effects Model which exploits within-group variation over time. Across-group variation is not used to estimate the regression coefficients, because this variation might reflect omitted variable bias. When using FE we assume that something within the individual may impact or bias the predictor or outcome variables and we need to control for this. This is the motivation behind the assumption of the correlation between entity’s error term and predictor variables. FE removes the effect of those time-invariant characteristics so we can assess the net effect of the predictors on the outcome variable. Moreover, time-invariant characteristics are unique to the individual and should not be correlated with other individual characteristics.

Publications:

- Nwakuya, M.T., Ijomah, M. A., (2017). "Fixed Effect Versus Random Effects Modeling in a Panel Data Analysis; A Consideration of Economic and Political Indicators in Six African Countries". International Journal of Statistics and Applications 2017, 7(6): 275-279

# Random Effects Regression for Panel Data

Another technique used to analyze Panel Data is Random Effects Model which exploits between-group variation over time. Random effects are estimated with partial pooling, while fixed effects are not. Partial pooling means that, if you have few data points in a group, the group's effect estimate will be based partially on the more abundant data from other groups. Random effects are simply the extension of the partial pooling technique as a general-purpose statistical model. This enables principled application of the idea to a wide variety of situations, including multiple predictors, mixed continuous and categorical variables, and complex correlation structures. The main assumption made in RE is that the deviations of subgroups from a parent follow the distribution of a random variable.

Publications:

- Gurka, M. & Kelley, G. and Edwards, L., (2012)." Fixed and random effects models". Wiley Interdisciplinary Reviews: Computational Statistics. 4(10)

# Hausman-Wu Test

The (Durbin-Wu-)Hausman (1978) test (also called as the Hausman specification test)in general detect endogenous regressors (explainatory variables) in regression model. Endogenous variables have values that are determined by other variables in the system. In panel data analysis, the Hausman test can help to choose between FE model and RE model. Under the null hypothesis the prefered model is RE model and under the alternative hypothesis the model FE is preferred. This test looks to see if there is a correlation between the unique (time-invariant) erros and the regressors in the model. The null hypothesis is that there is no correlation between the two.

Publications:

- Bell, A., Fairbrother, M. and Jones, K., (2018). "Fixed and random effects models: making an informed choice". Qual Quant. 53(2019):1051–1074

# Heckman-2 Step

The Heckman-2 Step method (often referred as Heckman Correction) is a statistical technique to solve the selection bias of in the data. correct bias from non-randomly selected samples, or to correct the incidentally truncated dependent variables. This is achieved by explicitly modelling the individual sampling probability of each observation (Step 1: Selection equation) together with the conditional expectation of the dependent variable (Step 2: Outcome equation). The resulting likelihood function is mathematically similar to the Tobit model for censored dependent variables.

Publications:

- Heckman, J. (1979). "Sample Selection Bias as a Specification Error". Econometrica. 47 (1): 153–61

Data Preparation for Regression Analysis

File includes the following operations:

- Generating variable

- Generating categorical variable

- Generating dummy variable

- Generating interaction variable

- Rename a variable

- Showing all possible values of the variable

- Drop a variable

- Drop all the observations with missing value

- Give a label to a variable that describes the variable

**Data and Results**

Summative Analysis of The Data

Summary Statistics of the Data

. summarize

Variable | Obs Mean Std. Dev. Min Max

-------------+---------------------------------------------------------

mbreastt | 2,756 3.347968 2.162987 1 7

cmkilo | 4,714 3.359744 .5864993 .5681818 5.511364

cmbweight | 5,050 2.02099 .3759915 1 3

ageyrs7 | 5,050 7.219053 .2442148 6.53 8.23

agemnths7 | 5,050 86.68399 2.930204 78.45 98.82

-------------+---------------------------------------------------------

reading7 | 5,050 47.18317 18.25874 -1 89

patconstr7 | 5,050 19.06832 6.992164 -1 63

maths7 | 5,050 9.796832 2.825889 -1 15

adnoba00 | 5,050 1 0 1 1

equivinc1 | 5,050 323.3627 204.0528 -1 1282.46

-------------+---------------------------------------------------------

maocc1 | 5,050 1.243168 2.557533 -1 7

paocc1 | 5,050 2.309901 2.646238 -1 7

chcnuma0 | 5,050 1 0 1 1

sex | 5,050 1.504752 .5000269 1 2

housing | 5,050 2.257426 .948187 -8 6

-------------+---------------------------------------------------------

country1 | 5,050 1 0 1 1

lessmedinc1 | 5,050 .2718812 .4480779 -1 1

maquals | 5,050 16.90158 33.23204 -1 96

paquals | 5,050 13.75168 30.78416 -1 96

pse5 | 5,050 21.39723 4.087893 0 27

-------------+---------------------------------------------------------

lct | 5,049 6.885126 1.669466 0 9

r | 5,050 6.578218 1.721188 0 9

w | 5,050 6.025743 1.977926 0 9

cll5 | 5,050 25.83069 6.686072 0 36

nlc | 5,050 7.388911 1.374598 0 9

-------------+---------------------------------------------------------

maths5 | 5,050 20.77109 4.372382 0 27

kuw5 | 5,050 6.810495 1.557484 0 9

pd5 | 5,050 7.329703 1.324095 0 9

cd5 | 5,049 6.772826 1.467316 0 9

fsptotal5 | 5,050 88.91069 17.38234 0 117

-------------+---------------------------------------------------------

intdate14 | 5,050 5.546535 2.381581 1 12

mob | 5,050 6.612673 3.427615 1 12

yob | 5,050 2000.651 .4767357 2000 2001

testage14 | 5,050 13.76673 .4501735 13 15

ethnincity | 5,050 1.691683 2.074229 -9 11

-------------+---------------------------------------------------------

BMI14 | 5,050 19.98032 6.622415 -1 46.49

obesity14 | 5,050 2.297822 1.149205 -1 4

wordtest14 | 5,050 6.607327 3.25421 -3 19

behaviour14 | 5,050 7.744356 5.914238 -1 38

friends7 | 5,050 1.397228 1.145847 -9 3

-------------+---------------------------------------------------------

worried7 | 5,050 2.032475 1.625253 -9 3

sad7 | 5,050 2.053861 1.429214 -9 3

bullied7 | 5,050 2.264752 1.491975 -9 3

momyob | 5,050 1970.733 5.648807 1944 1987

momhappy3 | 5,050 4.24198 3.116803 -1 8

-------------+---------------------------------------------------------

dog3 | 5,050 .1532673 .3721784 -1 1

childcare | 5,050 2.230891 1.611431 1 6

snacks5 | 5,050 5.956436 14.40445 1 96

Regressive Analysis of the Data with cmkilo as the Subject Point of Reference

. regress cmkilo ageyrs7 agemnths7 reading7 equivinc1

Source | SS df MS Number of obs = 4,714

-------------+---------------------------------- F(4, 4709) = 14.49

Model | 19.712457 4 4.92811426 Prob > F = 0.0000

Residual | 1601.47228 4,709 .340087552 R-squared = 0.0122

-------------+---------------------------------- Adj R-squared = 0.0113

Total | 1621.18474 4,713 .343981485 Root MSE = .58317

------------------------------------------------------------------------------

cmkilo | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

ageyrs7 | -.048072 2.986263 -0.02 0.987 -5.902545 5.806401

agemnths7 | .0007142 .248875 0.00 0.998 -.4871972 .4886256

reading7 | .0018935 .00048 3.94 0.000 .0009524 .0028346

equivinc1 | .0002287 .0000429 5.33 0.000 .0001445 .0003129

\_cons | 3.481363 .2528974 13.77 0.000 2.985566 3.977161

------------------------------------------------------------------------------

Data Regression with breast as the subject variable

. . regress mbreastt ageyrs7 agemnths7 reading7 equivinc1

Source | SS df MS Number of obs = 2,756

-------------+---------------------------------- F(4, 2751) = 25.19

Model | 455.419751 4 113.854938 Prob > F = 0.0000

Residual | 12433.8789 2,751 4.51976695 R-squared = 0.0353

-------------+---------------------------------- Adj R-squared = 0.0339

Total | 12889.2986 2,755 4.6785113 Root MSE = 2.126

------------------------------------------------------------------------------

mbreastt | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

ageyrs7 | -10.32548 14.13124 -0.73 0.465 -38.0344 17.38343

agemnths7 | .8659638 1.177622 0.74 0.462 -1.443149 3.175076

reading7 | .0101949 .0023152 4.40 0.000 .0056552 .0147346

equivinc1 | .0015357 .0001967 7.81 0.000 .00115 .0019215

\_cons | 1.792251 1.190838 1.51 0.132 -.5427769 4.127278

------------------------------------------------------------------------------

Multivariate Factor Analysis of Both Subjects

. mvreg mbreastt cmkilo = ageyrs7 agemnths7 reading7 patconstr7 equivinc1 pse5 lct r w cll5 nlc cd5 pd5 kuw5 maths5 nlc

note: nlc omitted because of collinearity

Equation Obs Parms RMSE "R-sq" F P

--------------------------------------------------------------------------

mbreastt 2,576 16 2.12075 0.0463 8.294043 0.0000

cmkilo 2,576 16 .5863073 0.0432 7.70491 0.0000

------------------------------------------------------------------------------

| Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

mbreastt |

ageyrs7 | -10.55264 14.58108 -0.72 0.469 -39.14455 18.03927

agemnths7 | .878604 1.214993 0.72 0.470 -1.503865 3.261073

reading7 | .0051958 .0029719 1.75 0.081 -.0006318 .0110233

patconstr7 | .0218499 .0065805 3.32 0.001 .0089463 .0347535

equivinc1 | .001324 .0002073 6.39 0.000 .0009174 .0017306

pse5 | .0186756 .0184831 1.01 0.312 -.0175678 .0549189

lct | -.0933591 .0660681 -1.41 0.158 -.2229114 .0361933

r | .0710861 .0750169 0.95 0.343 -.0760138 .2181861

w | -.0702521 .0751472 -0.93 0.350 -.2176076 .0771034

cll5 | .0146872 .0443349 0.33 0.740 -.0722487 .1016231

nlc | .0565378 .0690957 0.82 0.413 -.0789513 .1920269

cd5 | -.0031262 .0503351 -0.06 0.950 -.1018279 .0955754

pd5 | -.0685964 .0499802 -1.37 0.170 -.1666021 .0294092

kuw5 | .0921075 .0499005 1.85 0.065 -.005742 .189957

maths5 | -.004538 .0285877 -0.16 0.874 -.0605953 .0515193

nlc | 0 (omitted)

\_cons | 1.60928 1.257416 1.28 0.201 -.8563759 4.074936

-------------+----------------------------------------------------------------

cmkilo |

ageyrs7 | -.7196932 4.031118 -0.18 0.858 -8.624276 7.18489

agemnths7 | .0554155 .3358997 0.16 0.869 -.6032473 .7140782

reading7 | .0001557 .0008216 0.19 0.850 -.0014554 .0017668

patconstr7 | .0098336 .0018193 5.41 0.000 .0062662 .0134009

equivinc1 | .0001588 .0000573 2.77 0.006 .0000464 .0002712

pse5 | -.0024861 .0051099 -0.49 0.627 -.012506 .0075338

lct | .0002691 .0182653 0.01 0.988 -.0355473 .0360854

r | .0158846 .0207393 0.77 0.444 -.024783 .0565522

w | -.0053379 .0207754 -0.26 0.797 -.0460762 .0354003

cll5 | .0012145 .0122569 0.10 0.921 -.02282 .025249

nlc | .0164915 .0191024 0.86 0.388 -.0209661 .0539492

cd5 | -.0070477 .0139158 -0.51 0.613 -.0343349 .0202396

pd5 | -.002212 .0138176 -0.16 0.873 -.0293069 .0248829

kuw5 | .0289751 .0137956 2.10 0.036 .0019234 .0560268

maths5 | .0009013 .0079034 0.11 0.909 -.0145964 .016399

nlc | 0 (omitted)

\_cons | 3.167239 .3476281 9.11 0.000 2.485578 3.8489

------------------------------------------------------------------------------

The OLS estimates show that mt5-mt8 are not statistically significant at 90% significance level. In other words, the model predicted that mothers who worked as small employers, lower supervisory& technical, semi-routine and routine have no relationship with children’s mathematical performance. While, mothers work as higher managerial& professor, lower managerial& professor and intermediate are predicted to have positive impact on mathematical development of children at age 5. All of fathers’ occupation levels show a statistically significant relationship with mathematical development at 90% significance level. Comparing the coefficients of all the occupation categories, we found mothers who have lower managerial jobs or are professors (mt3) have the greatest positive impact on children’s mathematical development, the coefficient of which is 1.0369. That indicates children with mothers worked as lower managers and professors will have a 1.0369 higher score than children with mothers who have never worked or are long term unemployment. And fathers who work as higher manager and professors (dd2) have the greatest positive influence on children’s mathematical development. In other words, when fathers are higher manager and professors, children will perform a score of 1.9438 higher than children with fathers who have never worked or are long term unemployment when holding other variables constant. In terms of the influence of marital status, the model estimated it was statistically significant at 95% level with probability of 0.043. The coefficient is -0.3413 which implies the children who live in a single parent family will have a score of 0.3413 lower than the children who live with both parents. In reality, it implies the education from both parents’ sides is crucial for promoting children’s mathematical development.

**Conclusion**

The OLS estimates of our original unrestricted model indicated several variables were statistically insignificant, including mother’s smoke status, mothers’ occupations who are smaller employers, lower supervisory& technical, semi-routine and routine. It may indicate that these independent variables cannot explain the change in children’s mathematical performances. Looking at the goodness-of-fit in this model, adjusted R-squared indicates that 13.09% of the response variable variation can be explained by the multiple linear regression model. The value is not high while we still drew some statistically significant predictors, the reason probably is it is hard to interpret human’s behaviour. The constant of 16.2453 indicates that when the children are boys living with both parents, mothers who never smoked, parents are never educated and parents have never worked or are long term unemployed, the boys’ mathematical development score is 16.2453. The coefficient of sex variable is 1.328613 with a probability of 0, which means the influence of gender on children’s mathematical development is statistically significant. And girls are estimated to performance 1.3286 units better than boys holding other variables constant**.**

**Reference**

Breslow, N. E. & Day, N.E.(1980). Statistical Methods in Cancer Research: Vol 1-The Analysis of Case-Control Studies, International Agency for Research on Cancer: Lyon, pp. 1-40, 73-78, 84-115, 122-157, 280-289, 349-351.

Breslow, N. E. & Day, N.E.(1987). Statistical Methods in Cancer Research: Vol II-The Design and Analysis of Cohort Studies. International Agency for Research on Cancer: Lyon, pp.21, 65, 108-109, 336-344, 363-365.

Ingelfinger, J.A., Mosteller, F., Thibodeau, L.A. & Ware, J.H. (1994). Biostatistics in Clinical Medicine. 3rd ed. New York: McGraw Hill, pp. 323-328.

Mausner, J.S. & Kramer, S. (1985). Epidemiology: An Introductory Text. Philadelphia: W.B. Saunders Co., pp.43-64, 312-323.

Selvin, S. (1996). Statistical Analysis of Epidemiologic Data. New York: Oxford University Press, pp. 36, 93-96. STATA Reference Manual Release 7 Reference A-G., College Station Texas: Stata Press, pp. 455, 466.

**Raw Codebook Analysis of Data**

. codebook

---------------------------------------------------------------------------------------------------------------------------

mbreastt Length of breastfeeding period in months

---------------------------------------------------------------------------------------------------------------------------

type: numeric (float)

range: [1,7] units: 1

unique values: 7 missing .: 2,294/5,050

tabulation: Freq. Value

829 1

402 2

357 3

327 4

213 5

246 6

382 7

2,294 .

---------------------------------------------------------------------------------------------------------------------------

cmkilo Cohort member's birthweight

---------------------------------------------------------------------------------------------------------------------------

type: numeric (float)

range: [.56818181,5.5113635] units: 1.000e-08

unique values: 438 missing .: 336/5,050

mean: 3.35974

std. dev: .586499

percentiles: 10% 25% 50% 75% 90%

2.65 3.01136 3.38068 3.75 4.03409

---------------------------------------------------------------------------------------------------------------------------

cmbweight Categorical variable for birth weight of child

---------------------------------------------------------------------------------------------------------------------------

type: numeric (float)

label: cmbweight

range: [1,3] units: 1

unique values: 3 missing .: 0/5,050

tabulation: Freq. Numeric Label

305 1 Low

4,334 2 Normal

411 3 High

---------------------------------------------------------------------------------------------------------------------------

ageyrs7 S4 CM Age at interview in years

---------------------------------------------------------------------------------------------------------------------------

type: numeric (double)

label: DCAGEY00, but 137 nonmissing values are not labeled

range: [6.53,8.23] units: .01

unique values: 137 missing .: 0/5,050

examples: 7.01

7.14

7.28

7.44

---------------------------------------------------------------------------------------------------------------------------

agemnths7 S4 CM Age at interview in months

---------------------------------------------------------------------------------------------------------------------------

type: numeric (double)

label: DCAGEM00, but 447 nonmissing values are not labeled

range: [78.45,98.82] units: .01

unique values: 447 missing .: 0/5,050

examples: 84.13

85.78

87.45

89.33

---------------------------------------------------------------------------------------------------------------------------

reading7 S4 CM Total score for Word Reading test.

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: DCWRSC00, but 90 nonmissing values are not labeled

range: [-1,89] units: 1

unique values: 91 missing .: 0/5,050

examples: 31

45

54

64

---------------------------------------------------------------------------------------------------------------------------

patconstr7 S4 CM Pattern Construction Total Raw Score

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: DCTOTS00, but 45 nonmissing values are not labeled

range: [-1,63] units: 1

unique values: 46 missing .: 0/5,050

examples: 13

17

21

25

---------------------------------------------------------------------------------------------------------------------------

maths7 S4 CM Maths Test Score (Total Raw Score)

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: MTOTSCOR, but 16 nonmissing values are not labeled

range: [-1,15] units: 1

unique values: 17 missing .: 0/5,050

examples: 8

9

11

12

---------------------------------------------------------------------------------------------------------------------------

adnoba00 S1 DV Number of CMs in household

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: ADNOBA00

range: [1,1] units: 1

unique values: 1 missing .: 0/5,050

tabulation: Freq. Numeric Label

5,050 1 One baby

---------------------------------------------------------------------------------------------------------------------------

equivinc1 S1 DV OECD equivalised income

---------------------------------------------------------------------------------------------------------------------------

type: numeric (double)

label: ADOEDE00, but 4292 nonmissing values are not labeled

range: [-1,1282.46] units: .01

unique values: 4,293 missing .: 0/5,050

examples: 146.11

236.38

329.54

463.92

---------------------------------------------------------------------------------------------------------------------------

maocc1 S1 MAIN Respondent NS-SEC 7 classes (current job)

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: AMD07S00

range: [-1,7] units: 1

unique values: 8 missing .: 0/5,050

tabulation: Freq. Numeric Label

2,452 -1 Not applicable

273 1 Hi manag/prof

895 2 Lo manag/prof

576 3 Intermediate

131 4 Small emp and s-emp

108 5 Low sup and tech

430 6 Semi routine

185 7 Routine

---------------------------------------------------------------------------------------------------------------------------

paocc1 S1 PARTNER Respondent NS-SEC 7 classes (current job)

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: APD07S00

range: [-1,7] units: 1

unique values: 8 missing .: 0/5,050

tabulation: Freq. Numeric Label

1,336 -1 Not applicable

736 1 Hi manag/prof

977 2 Lo manag/prof

194 3 Intermediate

484 4 Small emp and s-emp

544 5 Low sup and tech

380 6 Semi routine

399 7 Routine

---------------------------------------------------------------------------------------------------------------------------

chcnuma0 S3 CM Cohort Member Number C1

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: LABA, but 1 nonmissing value is not labeled

range: [1,1] units: 1

unique values: 1 missing .: 0/5,050

tabulation: Freq. Numeric Label

5,050 1

---------------------------------------------------------------------------------------------------------------------------

sex S3 CM Cohort Member Sex C1

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: LABB

range: [1,2] units: 1

unique values: 2 missing .: 0/5,050

tabulation: Freq. Numeric Label

2,501 1 Male

2,549 2 Female

---------------------------------------------------------------------------------------------------------------------------

housing S3 HHQ Description of the selected dwelling

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: CHDTYP00

range: [-8,6] units: 1

unique values: 7 missing .: 0/5,050

tabulation: Freq. Numeric Label

3 -8 Don't Know

1,101 1 Detached house

1,921 2 Semi-detached house

1,713 3 Terraced house

239 4 Flat or maisonette - purpose

built

52 5 Flat or maisonnette - conversion

21 6 Other

---------------------------------------------------------------------------------------------------------------------------

country1 Country as at MCS1 sampling

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: ACOUNTRY

range: [1,1] units: 1

unique values: 1 missing .: 0/5,050

tabulation: Freq. Numeric Label

5,050 1 England

---------------------------------------------------------------------------------------------------------------------------

lessmedinc1 S1 DV OECD below 60% median poverty indicator

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: ADOEDP00

range: [-1,1] units: 1

unique values: 3 missing .: 0/5,050

tabulation: Freq. Numeric Label

7 -1 Missing data

3,663 0 Above 60% median

1,380 1 Below 60% median

---------------------------------------------------------------------------------------------------------------------------

maquals S1 MAIN Respondent NVQ highest level

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: AMDNVQ00

range: [-1,96] units: 1

unique values: 8 missing .: 0/5,050

tabulation: Freq. Numeric Label

1 -1 Not applicable

415 1 NVQ level 1

1,450 2 NVQ level 2

670 3 NVQ level 3

1,556 4 NVQ level 4

198 5 NVQ level 5

145 95 Overseas qual only

615 96 None of these

---------------------------------------------------------------------------------------------------------------------------

paquals S1 PARTNER Respondent NVQ highest level

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: APDNVQ00

range: [-1,96] units: 1

unique values: 8 missing .: 0/5,050

tabulation: Freq. Numeric Label

895 -1 Not applicable

282 1 NVQ level 1

1,062 2 NVQ level 2

619 3 NVQ level 3

1,322 4 NVQ level 4

248 5 NVQ level 5

162 95 Overseas qual only

460 96 None of these

---------------------------------------------------------------------------------------------------------------------------

pse5 s3 cm fsp: pse - personal, social and emotional development

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

range: [0,27] units: 1

unique values: 25 missing .: 0/5,050

mean: 21.3972

std. dev: 4.08789

percentiles: 10% 25% 50% 75% 90%

16 19 22 24 26

---------------------------------------------------------------------------------------------------------------------------

lct s3 cm fsp: lct - language for communication and thinking

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

range: [0,9] units: 1

unique values: 10 missing .: 1/5,050

mean: 6.88513

std. dev: 1.66947

percentiles: 10% 25% 50% 75% 90%

5 6 7 8 9

---------------------------------------------------------------------------------------------------------------------------

r s3 cm fsp: r - reading

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

range: [0,9] units: 1

unique values: 10 missing .: 0/5,050

mean: 6.57822

std. dev: 1.72119

percentiles: 10% 25% 50% 75% 90%

4 5 7 8 9

---------------------------------------------------------------------------------------------------------------------------

w s3 cm fsp: w - writing

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

range: [0,9] units: 1

unique values: 10 missing .: 0/5,050

mean: 6.02574

std. dev: 1.97793

percentiles: 10% 25% 50% 75% 90%

3 5 6 8 8

---------------------------------------------------------------------------------------------------------------------------

cll5 s3 cm fsp: cll - communication, language and literacy

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

range: [0,36] units: 1

unique values: 36 missing .: 0/5,050

mean: 25.8307

std. dev: 6.68607

percentiles: 10% 25% 50% 75% 90%

16 21 27 31 34

---------------------------------------------------------------------------------------------------------------------------

nlc s3 cm fsp: nlc - numbers as labels and for counting

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

range: [0,9] units: 1

unique values: 10 missing .: 0/5,050

mean: 7.38891

std. dev: 1.3746

percentiles: 10% 25% 50% 75% 90%

6 7 7 8 9

---------------------------------------------------------------------------------------------------------------------------

maths5 s3 cm fsp: md - mathematical development

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

range: [0,27] units: 1

unique values: 28 missing .: 0/5,050

mean: 20.7711

std. dev: 4.37238

percentiles: 10% 25% 50% 75% 90%

15 19 21 24 26

---------------------------------------------------------------------------------------------------------------------------

kuw5 s3 cm fsp: kuw - knowledge and understanding of the world

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

range: [0,9] units: 1

unique values: 10 missing .: 0/5,050

mean: 6.8105

std. dev: 1.55748

percentiles: 10% 25% 50% 75% 90%

5 6 7 8 8

---------------------------------------------------------------------------------------------------------------------------

pd5 s3 cm fsp: pd - physical development

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

range: [0,9] units: 1

unique values: 10 missing .: 0/5,050

mean: 7.3297

std. dev: 1.32409

percentiles: 10% 25% 50% 75% 90%

6 7 8 8 9

---------------------------------------------------------------------------------------------------------------------------

cd5 s3 cm fsp: cd - creative development

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

range: [0,9] units: 1

unique values: 10 missing .: 1/5,050

mean: 6.77283

std. dev: 1.46732

percentiles: 10% 25% 50% 75% 90%

5 6 7 8 8

---------------------------------------------------------------------------------------------------------------------------

fsptotal5 s3 cm fsp: total score

---------------------------------------------------------------------------------------------------------------------------

type: numeric (int)

range: [0,117] units: 1

unique values: 102 missing .: 0/5,050

mean: 88.9107

std. dev: 17.3823

percentiles: 10% 25% 50% 75% 90%

65 79 92 102 109

---------------------------------------------------------------------------------------------------------------------------

intdate14 Interview date (month)

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: FCINTM00

range: [1,12] units: 1

unique values: 12 missing .: 0/5,050

examples: 3 March

5 May

6 June

8 August

---------------------------------------------------------------------------------------------------------------------------

mob CM DOB (month)

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: FCCDBM00

range: [1,12] units: 1

unique values: 12 missing .: 0/5,050

examples: 3 March

5 May

8 August

10 October

---------------------------------------------------------------------------------------------------------------------------

yob CM DOB (year)

---------------------------------------------------------------------------------------------------------------------------

type: numeric (int)

label: FCCDBY00, but 2 nonmissing values are not labeled

range: [2000,2001] units: 1

unique values: 2 missing .: 0/5,050

tabulation: Freq. Numeric Label

1,763 2000

3,287 2001

---------------------------------------------------------------------------------------------------------------------------

testage14 CM Age last birthday

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: FCCAGE00, but 3 nonmissing values are not labeled

range: [13,15] units: 1

unique values: 3 missing .: 0/5,050

tabulation: Freq. Numeric Label

1,238 13

3,752 14

60 15

---------------------------------------------------------------------------------------------------------------------------

ethnincity S6 DV CM ethnic group classification - 11 categories

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: FDCE1100

range: [-9,11] units: 1

unique values: 14 missing .: 0/5,050

examples: 1 White

1 White

1 White

2 Mixed

---------------------------------------------------------------------------------------------------------------------------

BMI14 MCS6 Body Mass Index calculated (CLS)

---------------------------------------------------------------------------------------------------------------------------

type: numeric (double)

label: FCBMIN6, but 1472 nonmissing values are not labeled

range: [-1,46.49] units: .01

unique values: 1,473 missing .: 0/5,050

examples: 17.51

19.3

21.14

24

---------------------------------------------------------------------------------------------------------------------------

obesity14 MCS6 Obesity flag - UK90 thresholds

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: FCUK90O6

range: [-1,4] units: 1

unique values: 5 missing .: 0/5,050

tabulation: Freq. Numeric Label

301 -1 Not applicable

91 1 Underweight

3,069 2 Healthy weight

680 3 Overweight

909 4 Obese

---------------------------------------------------------------------------------------------------------------------------

wordtest14 CM Word activity score out of 20

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: FCWRDSC, but 20 nonmissing values are not labeled

range: [-3,19] units: 1

unique values: 22 missing .: 0/5,050

examples: 5

6

7

9

---------------------------------------------------------------------------------------------------------------------------

behaviour14 S6 DV Parent-reported CM SDQ Total Difficulties

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: FEBDTOT, but 35 nonmissing values are not labeled

range: [-1,38] units: 1

unique values: 36 missing .: 0/5,050

examples: 3

5

8

12

---------------------------------------------------------------------------------------------------------------------------

friends7 S4 CM How many friends do you have? (SCQ7)

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: DCSC0007

range: [-9,3] units: 1

unique values: 4 missing .: 0/5,050

tabulation: Freq. Numeric Label

40 -9 Not answered (9)

3,114 1 Lots

1,386 2 Some

510 3 Not many

---------------------------------------------------------------------------------------------------------------------------

worried7 S4 CM How often do you get worried? (SCQ12)

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: DCSC0012

range: [-9,3] units: 1

unique values: 4 missing .: 0/5,050

tabulation: Freq. Numeric Label

96 -9 Not answered (9)

253 1 All of the time

3,228 2 Some of the time

1,473 3 Never

---------------------------------------------------------------------------------------------------------------------------

sad7 S4 CM How often do you feel sad? (SCQ13)

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: DCSC0013

range: [-9,3] units: 1

unique values: 4 missing .: 0/5,050

tabulation: Freq. Numeric Label

74 -9 Not answered (9)

144 1 All of the time

3,602 2 Some of the time

1,230 3 Never

---------------------------------------------------------------------------------------------------------------------------

bullied7 S4 CM How often do other children bully you? (SCQ36)

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: DCSC0036

range: [-9,3] units: 1

unique values: 4 missing .: 0/5,050

tabulation: Freq. Numeric Label

71 -9 Not answered (9)

442 1 All of the time

1,977 2 Some of the time

2,560 3 Never

---------------------------------------------------------------------------------------------------------------------------

momyob S2 MAIN Person Date of Birth (year)

---------------------------------------------------------------------------------------------------------------------------

type: numeric (int)

label: LABT, but 41 nonmissing values are not labeled

range: [1944,1987] units: 1

unique values: 41 missing .: 0/5,050

examples: 1966

1969

1972

1976

---------------------------------------------------------------------------------------------------------------------------

momhappy3 S2 MAIN How happy are you

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: LABDI, but 7 nonmissing values are not labeled

range: [-1,8] units: 1

unique values: 9 missing .: 0/5,050

tabulation: Freq. Numeric Label

1,160 -1 Not applicable

88 1

91 2

128 3

229 4

582 5

1,322 6

1,430 7

20 8 Can't say

---------------------------------------------------------------------------------------------------------------------------

dog3 S2 MAIN Pets kept at home (Dog) MC1

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: LABAN

range: [-1,1] units: 1

unique values: 3 missing .: 0/5,050

tabulation: Freq. Numeric Label

22 -1 Not applicable

4,232 0 No

796 1 Yes

---------------------------------------------------------------------------------------------------------------------------

childcare S3 MAIN CM ever attended childcare providers C1 MC1

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: LABCV

range: [1,6] units: 1

unique values: 6 missing .: 0/5,050

tabulation: Freq. Numeric Label

2,577 1 Nursery School/Nursery Class

762 2 Playgroup

804 3 Pre-school

155 4 Childminder

379 5 Day Nursery (including

workplace/college creche)

373 6 None of these

---------------------------------------------------------------------------------------------------------------------------

snacks5 S3 MAIN: What CM eats between meals C1 MC1

---------------------------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: LABFL

range: [1,96] units: 1

unique values: 12 missing .: 0/5,050

examples: 1 Crisps and other similar snacks

4 Fruit (fresh, dried or tinned)

4 Fruit (fresh, dried or tinned)

4 Fruit (fresh, dried or tinned)

.