UK Numeracy Index

The UK Numeracy Index (UKNI) has been developed by National Numeracy and Experian. The index ranks areas of the UK by a predicted numeracy score. This can be used to help any local authority or organisation wanting to try to identify geographic areas of low numeracy as a first step to designing practical solutions to improve people's number skills and confidence within that area.

The UKNI is available as a percentile (1 to 100) for the following three geographic areas:

• Regions (12 areas);
• Local Authority Districts and Unitary Authorities (379 areas);
• Electoral Wards (8,869 areas).

Data sources

The UKNI has been created from numeracy data (confidence and skills) supplied by National Numeracy and four personal-level data variables from Experian.

The numeracy data was obtained from a sample of approximately 31,000 participants of the National Numeracy Challenge – a free website designed to help all adults to check and improve their numeracy.

Participants are asked to rate their own confidence with numbers (0 to 10) before completing a short check of their numeracy skills (0 to 20).

The modelling of the confidence and skills predictors used four of Experian's person-level variables:

• Gender;
• Affluence;
• Age;
• Employment Status.

National Numeracy is a charity which was established in 2012 to help raise low levels of numeracy among adults and children and to promote the importance of everyday number skills. Their aim is to challenge negative attitudes, influence public policy, and offer practical ways of helping adults and children improve their numeracy: in the community; the workplace; and formal education.

Methodology

Experian matched the numeracy data using the e-mail addresses supplied as part of signing up to the National Numeracy Challenge. Of the 31,000 records, c.6,400 (@20%) were successfully matched to Experian's demographics data.

Experian ran some initial profiling on the data which showed directional trends for certain demographics, including Age, Employment Status, and Gender. These then formed the backbone of the modelling and were used to enrich the matched records. The records were anonymised so that no one involved in the process could see the personal data.

Methodology (continued)

Confidence/skills scores were regressed on Age, Gender, Affluence, and Employment Status, with terms allowing for a curvilinear relation of the predictors to the response, and for interactions between the predictors (constrained by not being doubly non-linear). Weights were used to adjust for known biases in the demographic characteristics of those sampled compared to the general population.

Two numeracy models were created: one for confidence and one for skills. Plots were generated for each model to derive the importance of each predictor. For the confidence model, the predictor importance was (from highest to lowest):

- Gender;
- Affluence;
- Age;
- Employment Status.

While for the skills model, the predictor importance was (from highest to lowest):

- Affluence;
- Gender;
- Employment Status;
- Age.

An average confidence or skills score was calculated based on the number of households within each area. This allowed comparable areas of the UK to be ranked by percentile.

A single UKNI was then created for each area based on a combined view of the data from the confidence and skills models.

Predictor importance plots

The predicator importance plots show the relative contribution of each predictor (or the interaction between a pair of predictors) to the overall predictive power of the model for the confidence/skills score.
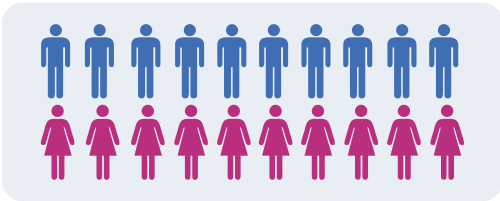

Confidence Predictors: Partial R^2

Skills Predictors: Partial R^2
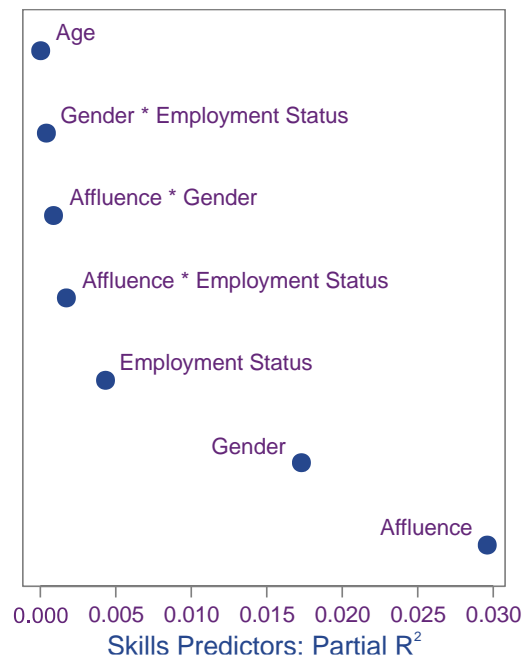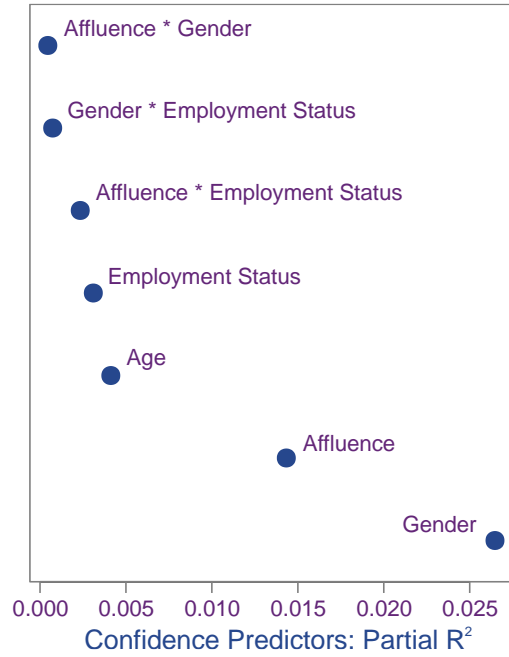
As can be seen, the plots confirm the relative importance of the 4 predictors that have been used to build the models.

UK Numeracy Index

## Experian's person-level variables

Gender

Experian's Gender variable identifies the gender of each individual living at each UK address.

A lookup of all known forenames is used to obtain the gender. If the forename lookup is inconclusive in establishing a definite male or female gender (because some names are ambiguous), or the forename is not found, then the person's gender is identified as unknown.

Affluence

Knowing a person's expected income is a valuable piece of information, but their overall affluence is a far more powerful indicator of their likely spending power. Experian's Affluence variable is used to identify an individual's affluence based on several key characteristics including income, property value, and other financial attributes.
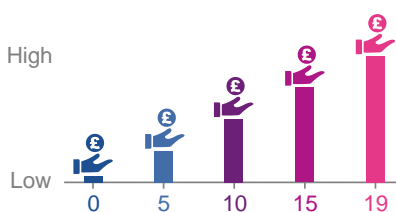
The input variables are converted to numeric values, and then used as predictive data within a gradient-boosted decision tree regression algorithm. The output score is ranked and divided into 20 bands (0 to 19). Band 0 represents the lowest affluence while band 19 the highest.

Age

Experian's Age variable identifies the actual age of each person. It is derived from Experian's Prospectable Date of Birth variable, if available, otherwise it is modelled.
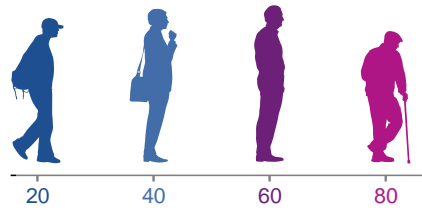
Employment Status

Experian's Employment Status variable identifies individual's employment status as six bands: full-time; student; unemployed; part-time; not working (e.g. homemakers or others voluntarily unemployed) and retired.
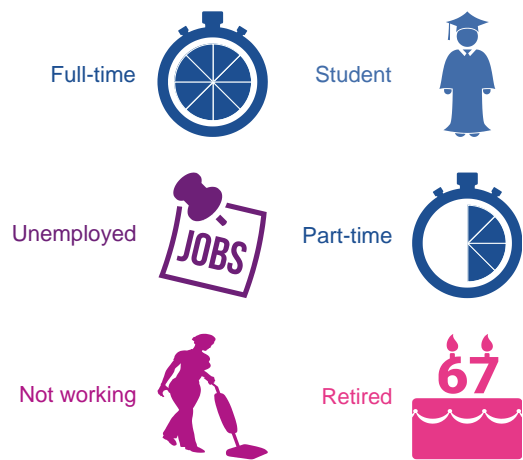
Employment Status is generated from YouGov sample data which is randomly split into development and test samples. Modelling techniques are used to select retired populations. Light-gradient boosting analysis is used to create probabilities for each status band and calibrated using local area targets by age bands, created using the Census Current Year Estimates.

Plots

The plots or charts in the following sections show the relationship between each predictor and the modelled confidence/skills score (holding all other predictors constant).
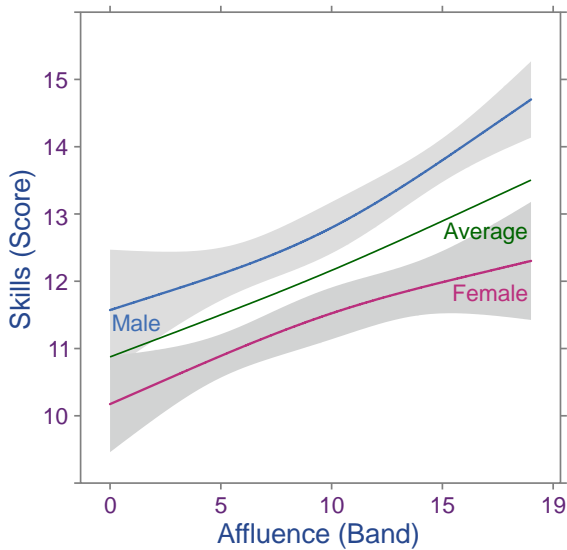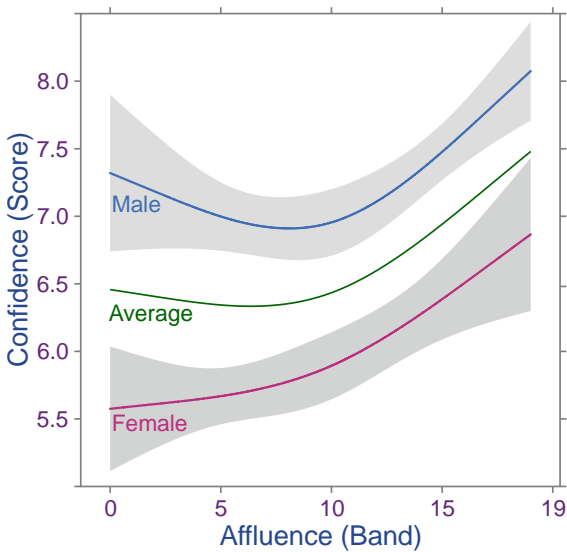
Affluence

For the two Affluence plots, blue (male) or pink (female) lines are used for each gender-specific categorical predictor. A green line is used to show the average score.

UK Numeracy Index

Affluence

The grey-shaded areas (for continuous predictors) and blue or pink lines give 95% confidence intervals for the modelled confidence/skills score. These grey areas tend to be wider at the extremes of continuous predictor values and for categories with fewer person counts.

The Affluence plots show that females generally have a lower score than males for every equivalent Affluence band. Both males and females see an increase in the skills score as affluence increases. This is also broadly true of confidence, although for younger males this is not so apparent.

Age

For the two Age plots, the definition of the blue, pink and green lines, and the grey-shaded areas are the same as those previously described for the Affluence plots opposite.

The Age plots show that females generally have a lower confidence and skills score than males for almost every equivalent age band.

At the very oldest ages we see a small amount of overlap of the grey bands. This small overlap is mainly because with fewer older people in the sample, the assessment of the difference cannot be as reliable or robust.

Both males and females see an increase in confidence score as age increases, although there no equivalent relationship for skills.
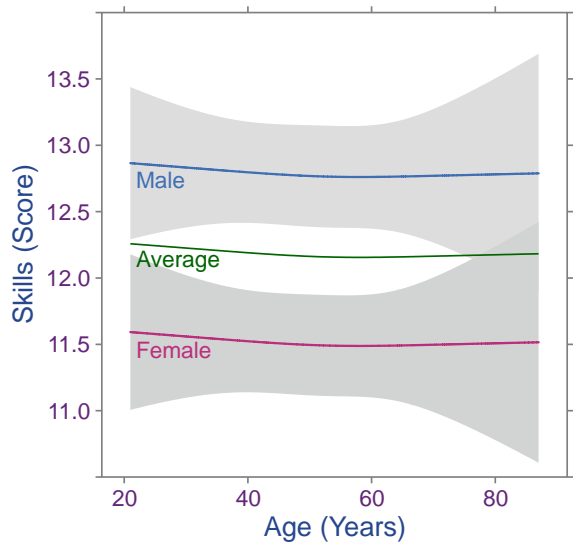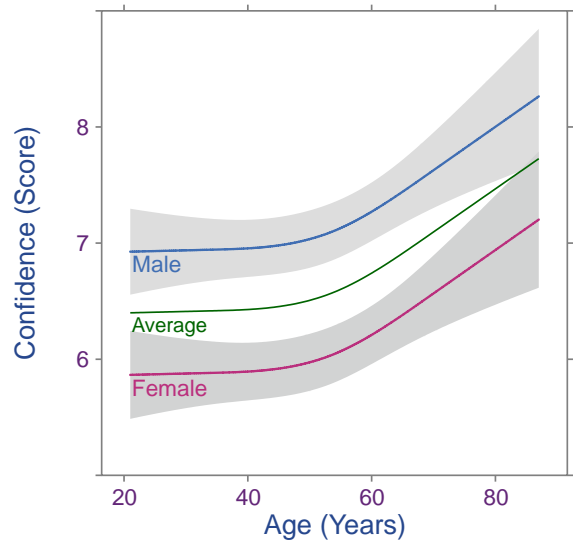
UK Numeracy Index

Employment Status

For the two Employment Status plots, blue (male) and pink (female) coloured dots represent the average numerical score for each Employment Status band, with blue lines showing the 95% confidence range. A green coloured dot is used to show the average score.

For the confidence score, the largest difference in the average is students, while full-time workers and retired show the smallest difference. Male students have the highest average confidence of any cohort shown.

For the skills score, the largest difference in the average is unemployed workers, while students and part-time workers show the smallest difference.

Both plots generally show that females have a lower score than males. Longer lines are shown for the unemployed, while full time and retired have shorter lines. This is mainly because with a relatively small number of unemployed people in the sample the assessment of the difference cannot be as reliable or robust.