LING82100: homework 8

(Due 5/13)

In this assignment you will fit a mixed-effects linear regression to reaction time data from the English Lexicon project (ELP; Balota et al. 2007), a lexical decision megastudy. The data consists of reaction times for roughly 180,000 trials from 251 subjects (with incorrect responses and non-word trials removed). You will prepare the data, fit a model, and provide a full interpretation for the results. The independent variables are:

1. `word`: target word,

2. `subjID`: subject ID,

3. `trial`: trial number,

4. `length`: the length of the word (in characters),

5. `OLD`: *orthographic Levenshtein distance* (Yarkoni et al. 2008), a measure of neighborhood density (cf. Coltheart's *n* in homework 4),

6. `sbtlx.freq`: word frequency in the SUBTLEX-US corpus (Brysbaert and New 2009), and

7. `sbtlx.basefreq`: frequency of the word's "base" (i.e, uninflected form) in SUBTLEX-US.

1 Data preparation

Load `elp.tsv`[1] and apply the following transformations.

- Log-transform RT.

- Convert subject ID `subjID` to a factor.[2]

- Center `trial`.

- Square `length` (as recommended by New et al. 2006), then standardize it.

- Standardize `OLD`.

- Add 1 to both `sbtlx.freq` and `sbtlx.basefreq`,[3] then log-transform and center them.

---

[1] `http://wellformedness.com/courses/LING82100/Data/elp.tsv`.

[2] Because it is numeric, R treats this like an integer, but it is a categorical variable.

[3] This is necessary in the case that any frequency is zero, because log 0 is undefined.

- If the Pearson correlation between the `sbltx.freq` and `sbtlx.basefreq` after the previous transformation is $> .5$, residualize `sbltx.basefreq` against `sbltx.freq`[4] and demonstrate that the resulting variables are uncorrelated.

What to turn in

Turn in R code showing the required transformations.

Hints

- Applying the transformations in the wrong order may give different results.

- You can either choose to overwrite the pre-transformed variables, or you can add new columns to your data frame; up to you.

Stretch goal

Determine whether there are any other non-trivial ($r > .5$) collinearities between the independent variables after the above set of transformations. If so, eliminate them by further residualization.

2   Modeling

Fit a mixed-effects logistic regression model on log-transformed RT with transformed `trial`, `length`, `OLD`, `sbtlx.freq`, and `sbtlx.basefreq` as fixed effects, and with `subjID` and `word` as random intercepts and interpret the results.

What to turn in

Turn in R code showing the model you fit. Report coefficients for all fixed effects, and apply the likelihood ratio test to all fixed effects.

Hints

- An sample table is shown in Table 1; simply fill in the elided values.

- Make sure to use the transformed variables and not the raw ones.

- Because each fixed effect is continously-valued, each likelihood ratio test should have d.f. $= 1$.

- Remember to round your p-values appropriately.

---

[4]That is, further transform `sbtlx.basefreq` so that it has no correlation with `sbtlx.freq`; you will not need to further transform `sbtlx.freq`.

	Coef.	S.E.	χ^2	$p(\chi^2)$
(Intercept)	...	...		
Trial number	...	...	...	...
Squared length	...	...	...	...
OLD	...	...	...	...
Word frequency	...	...	...	...
Base frequency	...	...	...	...

Table 1: Sample results table.

Stretch goal

Try to expand the model to also include a per-`subjID` random slope for `trial` in addition to the fixed effect of `trial`; does it work?

References

Balota, David A., Melvin J. Yap, Michael J. Cortese, Keith A. Hutchison, Brett Kessler, Bjorn Loftis, James H. Neely, Douglas L. Nelson, Greg B. Simpson, and Rebecca Treiman. 2007. The English Lexicon Project. *Behavior Research Methods* 39:445–459.

Brysbaert, Marc, and Boris New. 2009. Moving beyond Kučera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41:977–990.

New, Boris, Ludovic Ferrand, Christophe Pallier, and Marc Brysbaert. 2006. Re-examining the word length effect in word recognition: new evidence from the English Lexicon Project. *Psychonomic Bulletin and Review* 13:45–52.

Yarkoni, Tal, David A Balota, and Melvin J Yap. 2008. Moving beyond Coltheart's *n*: a new measure of orthographic similarity. *Psychonomic Bulletin and Review* 15:971–979.