9.S916, Assignment 3
due 21 April 2021

14 April 2021

1. **How bad is it when the assumption of normally distributed residuals is violated for linear regression?** For a simple linear regression model of the form $y = \alpha + \beta x + \epsilon$, replace the $\epsilon \sim N(0, \sigma)$ residual distribution with some other zero-centered non-normal residual distribution of your choice (natural alternatives might include gamma centered around zero, or some non-linear transformation of normal such as square root). Now, the Gauss–Markov theorem, which you can read about e.g. on StatLect, shows that least-squares linear regression (equivalently maximum-likelihood linear regression when the residuals are treated as normally distributed) still provides unbiased estimates of the regression weights α and β. What about the effect on statistical power?

   Consider the case where (i) the null is true, i.e. $\beta = 0$, and (ii) the null is false, i.e. $\beta \neq 0$. Using Monte Carlo simulation, assess the effect of underlyingly non-normal residuals that you are treating as normal in your regression analysis on Type I error rate (when $\beta = 0$) and statistical power (when $\beta \neq 0$) are affected. For the power comparison, you'll want to compare your simulations against an alternative situation where the errors are normally distributed with the same standard deviation as for the normal distribution. This latter case you can assess either with simulation or with standard power analyses for linear regression as in e.g. `R`'s `pwr` package.

2. **Interpreting effects from contrast schemes.** The dative alternation dataset of Bresnan et al. (2007) can be accessed by installing `R`'s `languageR` package, within which it named `dative`. For this dataset, fit a simple logistic regression model using as sole predictor the semantic class of the verb, using sum coding (accessible in `R` with `contr.sum()`) to numerically represent this five-level predictor with four numeric variables. The dependent variable here is `RealizationOfRecipient`, which takes on value `PP` for constructions like *I gave the apple to Susan* and `NP` for constructions like *I gave Susan the apple*. Use the likelihood ratio test for an **omnibus test** of the resulting model against an intercept-only null-hypothesis model (in `R` you can use `anova(model.null, model.alternative, test="Chisq")` for this). Then, from the fitted coefficients of the model, back out the predicted mean response for each of the five semantic classes, and confirm that they match the observed means in the dataset.

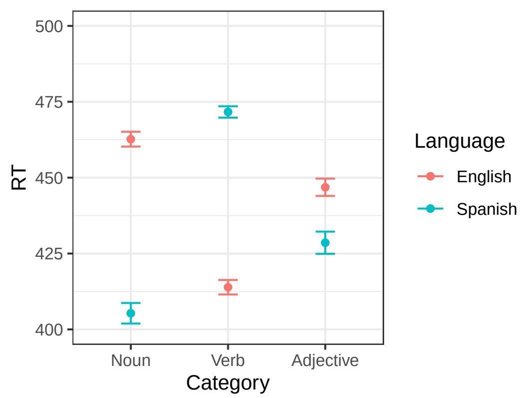

Figure 1: Condition means and standard errors for the hypothetical bilingual word recognition data

3. **Limits of the Wald z test statistic and robustness of the likelihood ratio test.** The Wald z statistic often breaks down for logistic regression when including a predictor allows (nearly) perfect prediction of the response variable. In contrast, the likelihood ratio test remains robust in these settings. Demonstrate this in simulations you develop yourself.

4. **Testing and interpreting main effects in the presence of an interaction.** This problem should be completed entirely with vanilla linear regression models (in `R`, the `lm()` function).

   Suppose you have a 3×2 experiment looking at how quickly English–Spanish bilinguals in a particular contact community recognize open-class words—nouns, adjectives, and verbs—in each of the two languages. You randomly choose twelve words in each class in each language. The below code generates hypothetical data for this experiment, which are summarized in Figure 1.

```
require(ggplot2,quietly=TRUE)
require(tidyverse,quietly=TRUE)
set.seed(1)
dat <- expand_grid(Category=factor(c("Noun","Verb","Adjective"),
                   levels=c("Noun","Verb","Adjective")),
                   Language=factor(c("English","Spanish")),
                   repetitions=1:12)
Beta <- matrix(400+c(60,15,45,5,70,25),3,2)
dat$RT <- with(dat,Beta[cbind(Category,Language)] + rnorm(nrow(dat),sd=10))
```

   There is manifestly a large interaction: average RTs for different categories differ across languages.

(a) **Task:** use nested model comparison (`R`'s `anova()` function) to conduct an F test to determine a p-value for the interaction.

However, there are also hints of main effects: averaging across the two languages, RTs seem to go in the order noun<adjective<verb, and RTs in Spanish look like they're a bit faster than RTs in English. The remainder of this problem is about how to evaluate these two main effects; I encourage you to consult Levy (2018) if you encounter difficulty.

(b) **Task:** try testing for the `Category` and `Language` main effects via a nested model comparison of the full main-effects model `RT ~ Category + Language` against the null models `RT ~ Language` and `RT ~ Category` respectively. Do the tests come out as statistically significant?

(c) **Task:** now try testing for each of these main effects *in the presence of the `Category:Language` interaction.* If you use `R` for this: because of how it handles factors (Chambers & Hastie, 1991; Levy, 2018), you will want to convert the factors into sets of numeric predictors and use those numeric predictors instead of the underlying factors in your regression models. Do the tests come out as statistically significant?

(d) **Task:** Explain the difference in the results arising from the two above approaches to testing the main effects.

(e) **Optional:** If you are familiar with traditional analysis of variance (ANOVA), conduct one on this dataset to simultaneously test the two main effects and the interaction. Do the results match either of the above results you obtained? Why? (Note: ANOVA is technically achieved with `R`'s `aov()` function, not `anova()`, which is for more general nested model comparison; the function names can be a bit misleading).

(f) **Task:** Interpret and critique the meaning of the two main effects and the interaction that you just tested. Do you think they are constructs of real potential scientific interest? Or are they an artifact of the experiment design?

References

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, H. (2007). Predicting the dative alternation. In G. Boume, I. Kraemer, & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 69–95). Amsterdam: Royal Netherlands Academy of Science.

Chambers, J. M., & Hastie, T. J. (1991). Statistical models. In J. M. Chambers & T. J. Hastie (Eds.), *Statistical models in S* (pp. 13–44). Chapman; Hall.

Levy, R. (2018). *Using R formulae to test for main effects in the presence of higher-order interactions.*