LING82100: homework 4

(Due 3/18)

1 Power analysis

Anderson (1935) gives a table of measurements of the petals and sepals (the leafy enclosure around the petal) of several species of irises on Gaspé, Quebec. These are summarized in Table 1; there are $n = 50$ samples of each of the three species. You will perform power analysis on this data.

What to turn in

1. Compute Cohen's d for the unpaired two-sample comparison of the sepal width of *I. versicolor* and *I. virginica.* Is the effect size small, medium, or large?

2. Compute the power of the above test at $\alpha = .01$. That is, given that α-level, the sample sizes, and the Cohen's d you computed in the previous step, what is the power of this comparison?

3. Draw 50-element random samples of the *I. versicolor* and *I. virginica* sepal widths using the sample statistics in Table 1. Then perform the appropriate t-test and report your result.

Hints

- The formulae for the first portion are eq. 3–4 from last week's lecture notes.
- Use the `pwr` library for the second portion.
- Use `rnorm` for the third portion. Since this involves random sampling, you will obtain slightly different results each time unless you seed the random number generator using the `set.seed` function.

Stretch goal

Write a function `cohen.d` for computing the pooled sample standard deviation, with the following signature:

```
cohen.d <- function(xbar1, xbar2, s1, s2, n1, n2)
```

where `xbar1` and `xbar2` are the sample means, texttts1 and `s2` are the sample standard deviations, and `n1` and `n2` are the size of the two samples. Then, use it to (re)compute the answer above. You may want to consult documentation or tutorials about writing R functions before you begin.

	I. setosa	*I. versicolor*	*I. virginica*
Petal length	1.46 (0.17)	4.26 (0.47)	5.55 (0.55)
Petal width	0.25 (0.11)	1.33 (0.20)	2.03 (0.27)
Sepal length	5.01 (0.35)	5.94 (0.52)	6.59 (0.64)
Sepal width	3.43 (0.38)	2.77 (0.31)	2.97 (0.32)

Table 1: Sample mean (and standard deviation) for three species of *Iris*.

2 Correlation analysis

It has long been believed that speakers' judgments of the well-formedness of nonce words is conditioned in part by said words similarity to real words (Vitevitch and Luce 1998, 1999, Vitevitch et al. 1999); in particular, that nonce words situated in dense "neighborhoods"—i.e., those which are similar to existing English words— will be rated more well-formed. Albright and Hayes (2003) asked 20 English speakers to rate the well-formedness of roughly 90 nonce words on a 7-point scale where 1 was labeled "completely bizarre, impossible as an English word" and 7 was labeled "completely normal, would make a fine English word". The TSV file `albright_hayes.tsv`[1] contains, the Albright and Hayes ratings, aggregated across subjects, and Coltheart's N (Coltheart et al. 1977), a widely-used measure of *neighborhood density*, computed using the Irvine Phonotactic Online Dictionary (Vaden et al. 2009). You will perform correlation analyses to determine whether there is an association between well-formedness and neighborhood density.

What to turn in

- Inspect the two variables, using tables, summary statistics, histograms, and/or Q-Q plots. What, if anything, suggests that a non-parametric correlation statistic would be appropriate here?
- Perform the Spearman ρ rank correlation test and report the results.

Hints

- Use `cor` and/or `cor.test` for the second portion, noting that these apply the Pearson (**not** the Spearman) test by default.

Stretch goal

Make a scatterplot of the above data, with well-formedness rating on the y-axis.

References

Albright, Adam, and Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90:119–161.

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

Anderson, Edgar. 1935. The irises of the Gaspé Peninsula. *Bulletin of the American Iris Society* 59:2–5.

Coltheart, Max, Eddy J. Davelaar, Jon T. Jonasson, and Derek Besner. 1977. Access to the internal lexicon. In *Attention and performance VI*, ed. Stanislav Dornic, 535–555. Hillsdale, NJ: Lawrence Erlbaum.

Vaden, Kenneth, Harry R. Halpin, and Gregory S. Hickok. 2009. Irvine Phonotactic Online Dictionary. URL: `http://www.iphod.com`.

Vitevitch, Michael S., and Paul A. Luce. 1998. When words compete: levels of processing in perception of spoken words. *Psychological Science* 9:325–329.

Vitevitch, Michael S., and Paul A. Luce. 1999. Probabilistic phonotactics and neighborhood density in spoken word recognition. *Journal of Memory and Language* 40:374–408.

Vitevitch, Michael S., Paul A. Luce, David B. Pisoni, and Edward T. Auer. 1999. Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language* 68:306–311.