Stats in Ling Homework

Richard Littauer

February 21, 2012

- Given our current data, a tree ending in linear regression would be the best fit for a statistical test. As this is the case, it probably wouldn't be advantageous to run other tests on the whole data. For comparing the *wpm* differences between two levels of drinking, a *t*-test should suffice.
- See above.
- This data has been collated with the end goal of seeing what the average RT is for sentence completion in ditransitive cases using nonsense and sensical words. An example would be grammaticality judgements for the input: *I gave the girl* a ball / running. We want the mean of RT for grammatical and non-grammatical readings.
- Independent samples *t*-test:
    - Calculate the sample mean: 670ms.
    - Get the hypothesized true mean: 670ms *From slides.*
    - Calculate the standard deviation of the sampling distribution using this formula: $\sqrt{\frac{(value - mean\ value)^2}{N-1}}$
    = 5.47726
    - Calculate by subtracting the true mean from the sample mean, and dividing by the SD of the sampling distribution. *Due to the highly fabricated nature of the data, the true mean (670ms) minus the hypothesised mean (also 670ms) causes this to cancel out to zero. Which is highly significant, given the t-distribution table.*
    - Compare this number to the t-distribution table. P = essentially 0.

- Wilcoxon test: *Refer to hw5_data_wilcox for the appropriate table.*
    - Load in the dependant variable values
    - Mu is the expected median
    - Figure out the charge of x-mu
    - The x-mu value
    - Take the absolute value of x-mu
    - Figure out the rank
    - Figure out the rank with the charge of x-mu.

    See the hw5_data.csv file for details. The final *w* measure is 85.5, which is not statistically significant.
- Done. See R code for appropriate way to calculate t and Wilcoxon tests.