

Chapter 6: Exercises

- 1) Which linguistic disciplines use the notion of 'style' as one of their core concepts? How can it be operationalized?
- 2) Which of these cases of variation satisfy Labov's definition of a sociolinguistic variable and can thus be investigated using Labov's methods? Justify your answer.
 - a) h-dropping in different social contexts: i.e. the variation between the pronunciation of e.g. the word 'hair' as /heə/ or /eə/.
 - b) The variation between naming strategies for soft drinks in the US, e.g. *soda*, *pop* or *coke*. (<http://popvsoda.com/>)
 - c) The variation between the active and the passive construction, e.g. *I did it* vs. *It was done*.
 - d) The variation between verbs expressing preferences with a different intensity, e.g. *adore*, *love*, *like* and *don't mind*
 - e) The use of hedges in discourse and the variation between a hedged and unhedged utterance, e.g. *I sort of agree*. vs. *I agree*.
- 3) Look at the following ways of reporting the t-test, ANOVA and the Mann-Whitney U test. Identify erroneous or missing pieces of information.
 - a) $t = 2.77, p = .007, d = .69$.
 - b) $F(56) = 5.59, p < .00201359$.
 - c) $U = 705, p = 1.3$
- 4) Use the equations in this chapter to calculate the test statistics (i.e. not the p-values) for:
 - a) The use of *lovely* in male and female speech [t-test, Mann-Whitney U].
Male group ($n_1=10$): 0.91, 1.4, 2.18, 6.21, 2.63, 1.2, 0, 1.06, 6.49, 5.43
Female group ($n_2=10$): 8.84, 1.09, 12.47, 1.65, 3.93, 1.1, 4.11, 21.21, 2.51, 0.47
 - b) The use of *innit* in the speech of speakers from the South, Midlands and the North [one-way ANOVA].
South ($n_1 = 10$): 4.19, 29.29, 5, 30.43, 6.09, 12.77, 25.93, 0.61, 28.08, 15.94
Midlands ($n_2 = 10$): 9.68, 3.65, 1.2, 0, 2.07, 2.26, 5.18, 0, 0, 0
North ($n_3 = 10$): 9.09, 9.09, 0, 7.38, 0, 5.77, 0, 4.47, 0, 3.23
- 5) Calculate Cohen's d for the dataset a) in Question 4.
- 6) Use the Group comparison tool to check results of 4) and 5).
- 7) Interpret the following correspondence plot. It is based on BNC64, a corpus of informal British speech. 16 male and 16 female speakers are plotted in the graph according their use of different semantic types of certainty markers (*certainly*, *maybe*, *perhaps*, *possibly*, etc.) Each speaker is labelled according to their gender (F1a and M1a), number identifier (F1a) and sample number (F1a, F1b); there are two samples per speaker.

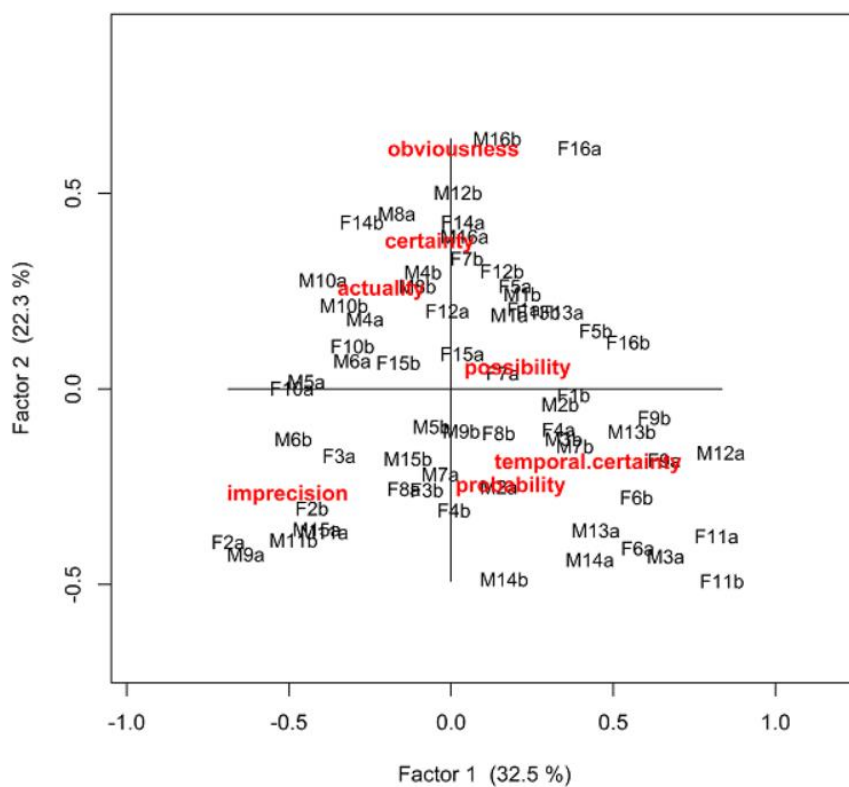


Figure 6.12 Correspondence analysis: Use of epistemic markers in BNC64

- 8) Interpret the following output from the mixed-effects model technique; the linguistic variable is the use of *must* competing with other strong modal terms such as *have to* and *need to* in different genres of British and American English.

```

      AIC      BIC   logLik deviance df.resid
2412.3   2456.9  -1198.1  2396.3    1943

Scaled residuals:
   Min       1Q   Median       3Q      Max
-1.9037 -0.6202 -0.3895  0.7452  2.4673

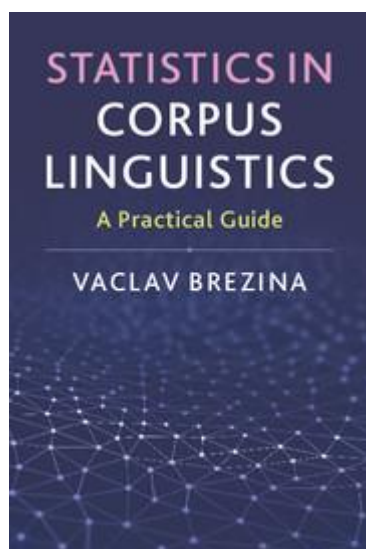
Random effects:
 Groups Name   Variance Std.Dev.
Text   (Intercept) 1.52    1.233
Number of obs: 1951, groups: Text, 743

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    0.1783    0.2364   0.754 0.450635
VarietyB_BR     0.1567    0.1515   1.034 0.300982
GenreB_Fiction -0.9825    0.2717  -3.616 0.000299 ***
GenreC_General -0.2894    0.2569  -1.126 0.259977
GenreD_Press   -0.7682    0.2920  -2.631 0.008511 **
SubjectB_I     -1.1864    0.2153  -5.511 3.58e-08 ***
SubjectC_you   -0.8857    0.2030  -4.363 1.29e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
    
```

- 9) *Do men swear more than women?* Use the BNC64 Search & Compare tool (corpora.lancs.ac.uk/bnc64) to test different socio-linguistic hypotheses about swearing in informal British speech. Pay attention to the different statistical measures and their interpretation.

Table 6.5 Swearing and gender: BNC64

Swear word	Statistically significant result?	Meaningful difference?
1.		
2.		
3.		
4.		



Brezina, V. (2018). [*Statistics in Corpus Linguistics: A Practical Guide*](#). Cambridge: Cambridge University Press.

Do you use language corpora in your research or study, but find that you struggle with statistics? This practical introduction will equip you to understand the key principles of statistical thinking and apply these concepts to your own research, without the need for prior statistical knowledge. The book gives step-by-step guidance through the process of statistical analysis and provides multiple examples of how statistical techniques can be used to analyse and visualise linguistic data. It also includes a useful selection of discussion questions and exercises which you can use to check your understanding.

The book comes with a Companion website, which provides additional materials (answers to exercises, datasets, advanced materials, teaching slides etc.) and [Lancaster Stats Tools online](#), a free click-and-analyse statistical tool for easy calculation of the statistical measures discussed in the book.