Linguistics 220: Quantitative Data Analysis Workshop
Calculating Variation
October 10, 2019

You are free and encouraged to use other methods/applications of calculating proportions and totals if you are comfortable. Below are instructions for making Pivot Tables in Google Sheets, one possibility for those of you that don't have experience doing this type of analysis.

Make a COPY of the dataset (either download as a spreadsheet or File > Make a copy in Google Drive). Please do not work directly from the original dataset!

Creating pivot tables in Google Sheets for raw counts:

1. Select the entire dataset for the tab you'll be working from, including the header row.
2. Go to Data > Pivot Table…
3. A new tab will open that includes your pivot table.
4. On the right side, a menu called "Pivot table editor" will pop up. This is where you will select the variable you're looking at, and find counts according to some social factor.
5. Under Rows click "Add." A menu will open that should contain the header for every column in the original data sheet. Select the variable you're looking at (e.g. Cot-Caught). This will create a row for every possible value in the column. For mergers, this should just be S and D, for morphosyntax, this should be Y, H, and N, etc. If you see multiple entries or entries that are unexpected, these are likely typos or spaces in the original data sheet. Use your judgment to decide which should be combined, if necessary.
6. Now, under Values, click "Add." Select the same variable you've entered under Rows. **Be sure to select Summarize by: COUNTA.** This will fill in the counts for each value in that column. The total should add up to the total number of speakers in the sample.
7. Now, you can add your social factor. Under Columns, click "Add," then select the social factor you want to investigate. Again, all of the possible values for this social factor will be filled in as separate columns (e.g. for Gender, you should see a column for F, M and N). This table now provides the counts for each variant, by social factor:

	A	B	C	D	E	F
1		**F**	**M**	**N**		**Grand Total**
2		0	0	0		0
3	**H**		8	8		16
4	**N**		81	60	2	143
5	**Y**		4	9		13
6	**Grand Total**	0	93	77	2	172
7						
8						

Calculating variation in the quantitative paradigm:

Remember, whatever method you use to obtain counts, to calculate usage in the quantitative paradigm, you need to calculate **proportions** for each social factor. A raw number doesn't mean much when we don't take into account the total numbers of speakers. You need to calculate the proportion of use for each variant, for each social group you're looking at.

For example, if you're looking at Cot-Caught by gender, you need to calculate: what proportion of people that have the merger (said "same") in each gender group (% merger in females, % merger in non-binary speakers, % merger in males). By comparing these proportions, you can see if there are differences in the merger between the gender groups. To obtain this, you'll calculate the following:

female *same* responses / total N females = % merger in females
male *same* responses / total N males = % merger in males
non-binary *same* responses / total N non-binary speakers = % merger in non-binary speakers

You can look at morphosyntactic variables similarly. You may choose to group the responses into two categories (would say, versus would not say), or look at all three options separately (would say, have heard, have never heard). Similarly, you may choose to group the frequency-based responses together for “lowkey” (for example, compare people who use “lowkey” at least once a week to those who use “lowkey” less than once a week). Again, always calculate a particular variant by a group of speakers as a proportion of the total number of speakers in that group. For “lowkey,” Kiesling (2004) has good examples of how to conduct this type of analysis in his discussion of “dude.”

Reporting quantitative results:

You will need to include one table per variable (five tables) and at least one visual chart total (at least one graph) in your paper. A visual comparison can make a trend very clear, while a table provides more complete data. In your table, you should include both the proportions you have calculated, by social factor, as well as the total numbers in each group. A good example is provided in Tagliamonte (2009), who looked at frequency of quotative verbs (verbs introducing a quotation) by time period (at right):

Table 6.1 Overall distribution of quotative verbs, university students

Quotatives	*British English 1997*		*Canadian English 1995*	
	%	N	%	N
say	31	209	36	219
go	18	120	22	135
be like	18	120	13	79
think	18	123	4	27
zero	10	66	20	123
Other	4	24	5	29
Total		665		612

In your graph, be sure to plot **proportions**. There are examples in the papers of plotting raw counts, but proportions are most informative in a visual comparison. You may use whatever type of chart you think best visualizes the data. One good example is provided below: (visualizing Trudgill’s (1974) ING data).

References

Kiesling, Scott. (2004) Dude. *American Speech* 79(3): 281-305.

Tagliamonte, Sali A. (2009) Be like: The new quotative in English. In Nikolas Coupland and Adam Jaworski (eds.) *The New Sociolinguistics Reader.* London: Palgrave MacMillan.

Trudgill, Peter. (1974) *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.