

How to read quantitative results

Adriana Picoral (adrianaps@arizona.edu)

August 31, 2022

Frequency tables

Variable totals

1. How many tokens total are there?
2. What is the variant of reference? (check Tagliamonte's textbook)
3. What is the population input?
4. Look at table 5.2 in Tagliamonte's book: what are the differences?

variable		n	percentage
	that expressed	395	0.21
	that omitted	1477	0.79
total	—	1,872.00	1.00

Frequency by factor group

1. How many levels are there in the recoded verb factor group?
2. What is the factor level with the highest “that” omission?

factor group levels		n	percent omitted	standard deviation
	Think	842	0.95	0.21
	Say	235	0.75	0.43
	Know	136	0.68	0.47
	OTHER	659	0.61	0.49
total	—	1,872.00	—	—

Frequencies don't tell the whole story. How do we know if there are actual statistical differences across groups?

Factor Weights

Factor Weights answer the following question:

- Based on the overall probability of omission, how does each factor level favor omission?

The overall probability of omission is 0.79.

1. Why are the Factor Weights different than the percentage of omission?

levels	n	%	fw
Think	842	0.95	0.84
Say	235	0.75	0.45
Know	136	0.68	0.35
OTHER	659	0.61	0.30

R output

More and more people are using R to analyze their data. Often, you see authors presenting their results like this:

```
##
## Call:
## glm(formula = dep_var_numeric ~ Verbs.1, family = "binomial",
##      data = compl_data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4788   0.3080   0.3080   0.8842   0.9868
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.7376     0.1833   4.024 5.72e-05 ***
## Verbs.1OTHER  -0.2710     0.2000  -1.355   0.175
## Verbs.1Say     0.3781     0.2377   1.591   0.112
## Verbs.1Think   2.2872     0.2459   9.300 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1929.2  on 1871  degrees of freedom
## Residual deviance: 1628.3  on 1868  degrees of freedom
## AIC: 1636.3
##
## Number of Fisher Scoring iterations: 5
```

Some important information can be retrieved from the model, like the variance explained, which in this case is 0.16 (McFadden's R-squared).

1. How do you interpret the results above?
2. How do you read the intercept?
3. Why is there a verb missing in the results?

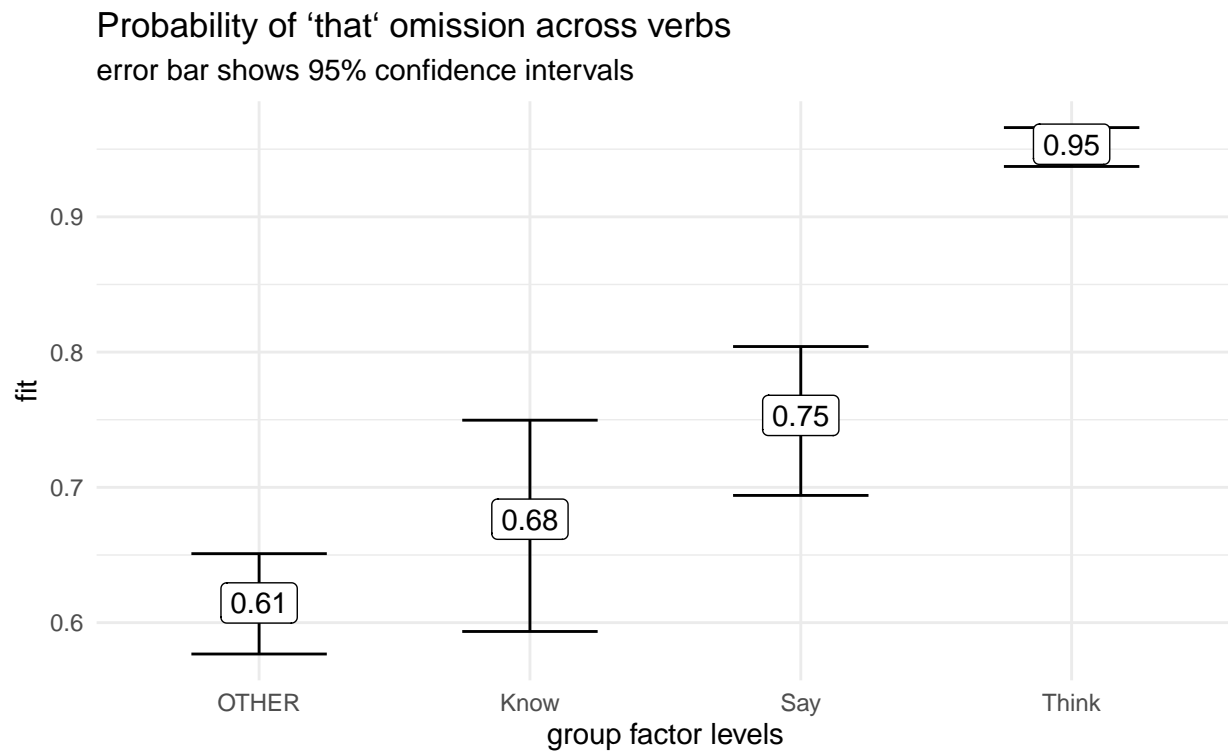
The raw output for regression is not ideal for language data. By default, the contrasts are set for experimental settings, which defines a control group as the intercept. That is useful when testing different types of drugs against a control group, but that's not what we do in LVC.

We don't have a control group for language data. Although the model is still the same, and it is producing the same results, the output is not formatted in a way that makes sense for LVC.

Here's the exact same results, from the same model, but presented in a way that makes more sense to humans:

factor group levels	fit	se	lower	upper
Think	0.95	0.01	0.94	0.97
Say	0.75	0.03	0.69	0.80
Know	0.68	0.04	0.59	0.75
OTHER	0.61	0.02	0.58	0.65

We can also visualize the results on a plot:



Rbrul output

1. What is shown in the output below?
2. What is the same and what is different from the previous tables?

```
model.basics
total.n df intercept input.prob grand.proportion
1872 4 1.336 0.792 0.789

model.fit
deviance AIC AICc Somers.Dxy R2
1628.318 1636.318 1636.34 0.495 0.304

Verbs.1
logodds n proportion factor.weight
Think 1.689 842 0.954 0.844
Say -0.220 235 0.753 0.445
Know -0.599 136 0.676 0.355
OTHER -0.870 659 0.615 0.295
```