Why We Need Better Statistics – A Case Study on eh in New Zealand English

Martin Schweinberger
www.martinschweinberger.de
University of Hamburg

16. Norddeutsches Linguistisches Kolloquium 2015
Leibniz-Universität Hannover

Research questions

- What happens if we use sub-optimal statistics?
- Is the use of discourse particles affected priming?
  Is *eh* affected by variables other than sociolinguistic factors (age, gender, etc.) or pragmatic functionality?

Table of Contents

Theoretical framework

Discourse particle *eh*

Priming

Variationist sociolinguistics

Discourse particle *eh*

- Syntactically optional, semantically bleached/empty discourse elements
- Use determined by discourse pragmatic functionality and sociolinguistic factors

(1) Speech Unit Final *eh* in NZE

a. oh we wont go that far *eh* (ICE-NZ:S1A-001#1:M)

b. its a really bad buzz *eh* (ICE-NZ:S1A-004#1:M)

c. yes so pete moves on and and clinton comes in and the girls will still stay here *eh* (ICE-NZ:S1A-037#1:M)

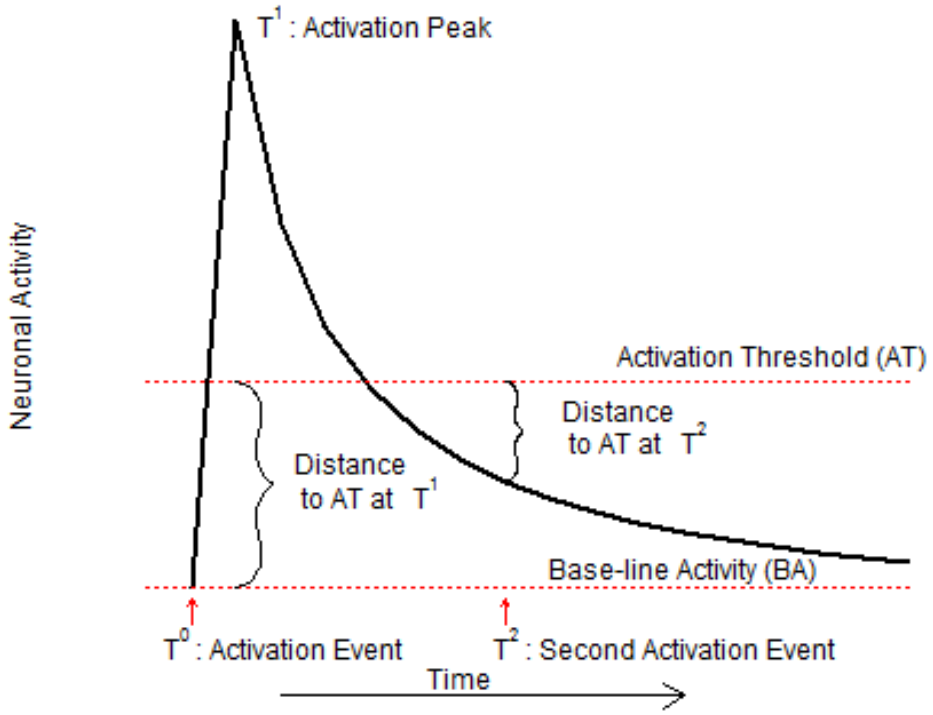
d. yeah yeah it is *eh* (ICE-NZ:S1A-047#1:A)

eh in New Zealand English (NZE): Properties

- Typically in turn-final position (non-polar tag).
- Stereotyped in NZE (vernacular NZE).
- Maori > Pakeha (British/European New Zealanders)
- In-group marker of ethnic identity (Meyerhoff 1994: 371)
- Men > women
- Maori men > (young) Pakeha women > Pakeha men > Maori women (interaction between ethnicity and gender) (cf. Meyerhoff 1994: 371)
- Young speakers > Old speakers

(Lexical) Priming

- Re-use of material that was used in previous utterances (cf. Tulving and Schacter 1990: 301)



Modern sociolinguistic theory

- Analysis of the relationship between linguistic variation and social factors (correlational/ variationist sociolinguistics)
- Very successful in finding and explaining systematic patterns that occur during language change (linguistic elements diffuse through the speech community)
- Shortcomings
  - Neglect of psycholinguistic factors (e.g. priming) (cf. Gries 2013)
  - Extensive use of fixed-effects regressions (Varbrul, GoldVarb) (Sankoff et al. 2005)
  - High α-error rate – model reports a factor as being significant although it is actually not (up to 80%, Johnson 2009: 369)

Methodological evolution of variationist sociolinguistics. . .

Traditional sociolinguistics	*Modern sociolinguistics*	*Tomorrow's sociolinguistics*
	Methodology	
Frequency analysis & bivariate statistics	Multivariate statistics	Sophisticated statistical modelling
(non-)parametric tests	GoldVarb analysis	R
χ^2-test, t-test, Wilcoxon Sign-Rank test	Logistic Regression	Generalized linear (mixed-effects) models, PCA, MDS
	Proponents	
Trudgill	Labov	Szmrecsanyi
Chambers	Rickford	?
Preston	Tagliamonte	?

Table of Contents

Data

- Most informal register (S1A: face-to-face conversation, telephone calls) of the New Zealand component of the *International Corpus of English* (ICE)

Data	Speakers (N)	Words (N)	Speech Units (N)	*eh* (N)
All spoken files	1,085	653,186	68,189	421
Only private dialogue	250	213,555	31,544	410
Only private dialogue with complete cases (Age, Gender, Ethnicity)	203	188,539	25,821	328

Table of Contents

Dependent variable		
suf.eh	nominal	yes no occurrence of speech unit final eh
Independent variable		
file.speaker.id	categorical	combined file and speaker id
age	nominal	old (40+) versus young (39-)
sex	nominal	male versus female
ethnicity	nominal	Pakeha versus Maori
priming	nominal	yes no occurrence within last 18 words before turn

Statistical design

- What is a Regression?
  - Statistical method
  - Measures the relationship between some phenomenon (dependent variable) and various predictors(independent variables/factors).
  - Example
    - Monthly income (dep.) $\sim$ degree of education (indep.)
    - Do people that have attended university earn, on average, more than people who have not attended university?

Statistical design

- What is a Regression?
  - Meaningful relationship $\rightarrow$ significant
  - Effect size $\rightarrow$ How strong is this relationship?
  - Example
    - A-levels (Abitur) correlates significantly with attending a university
    - Predictor (having A-levels) has a substantial effect size, i.e. it tells us that if you have your A-levels, then it is very likely that you will also have attended university.

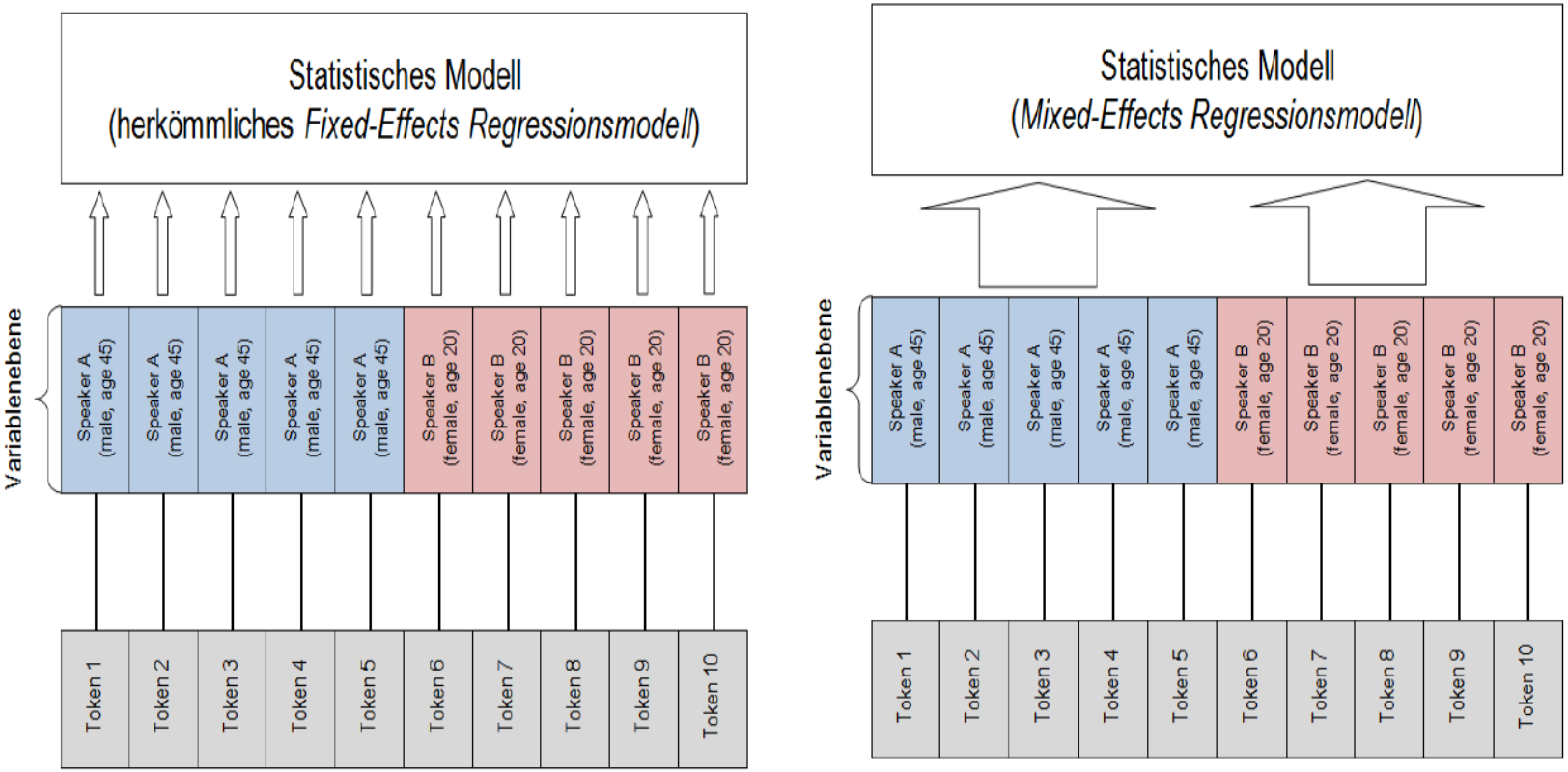
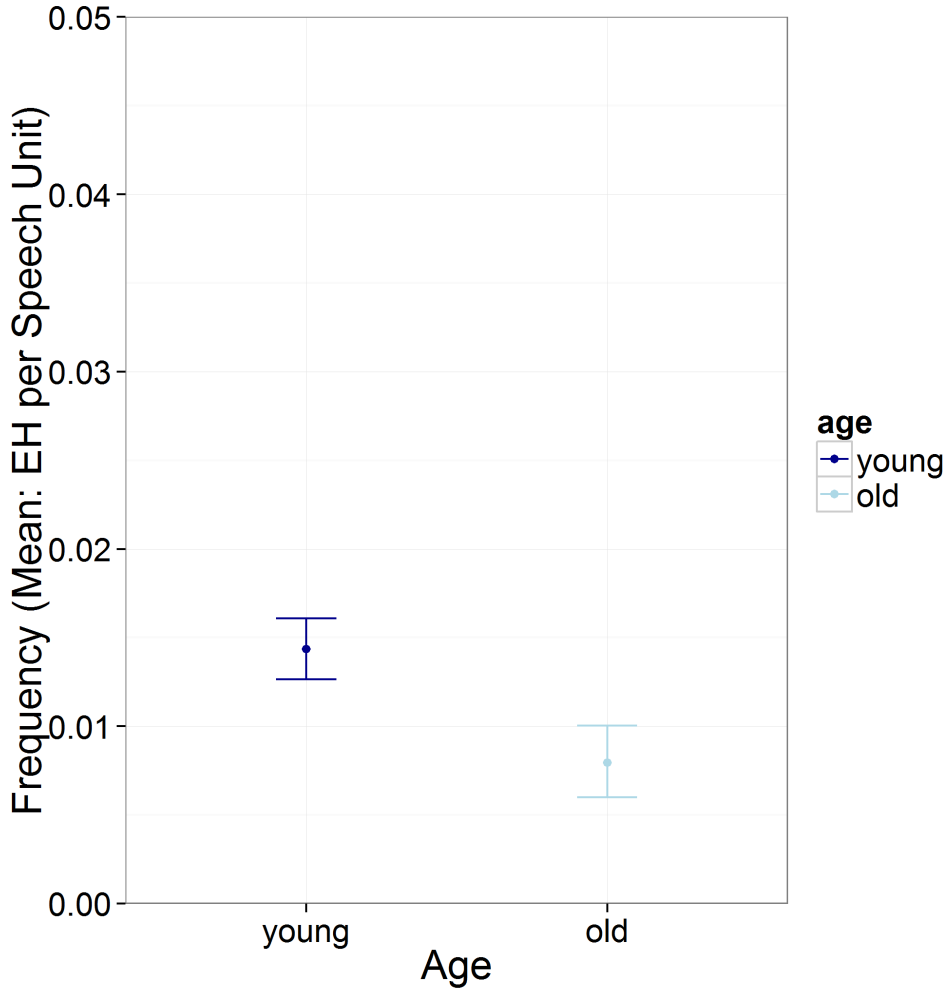


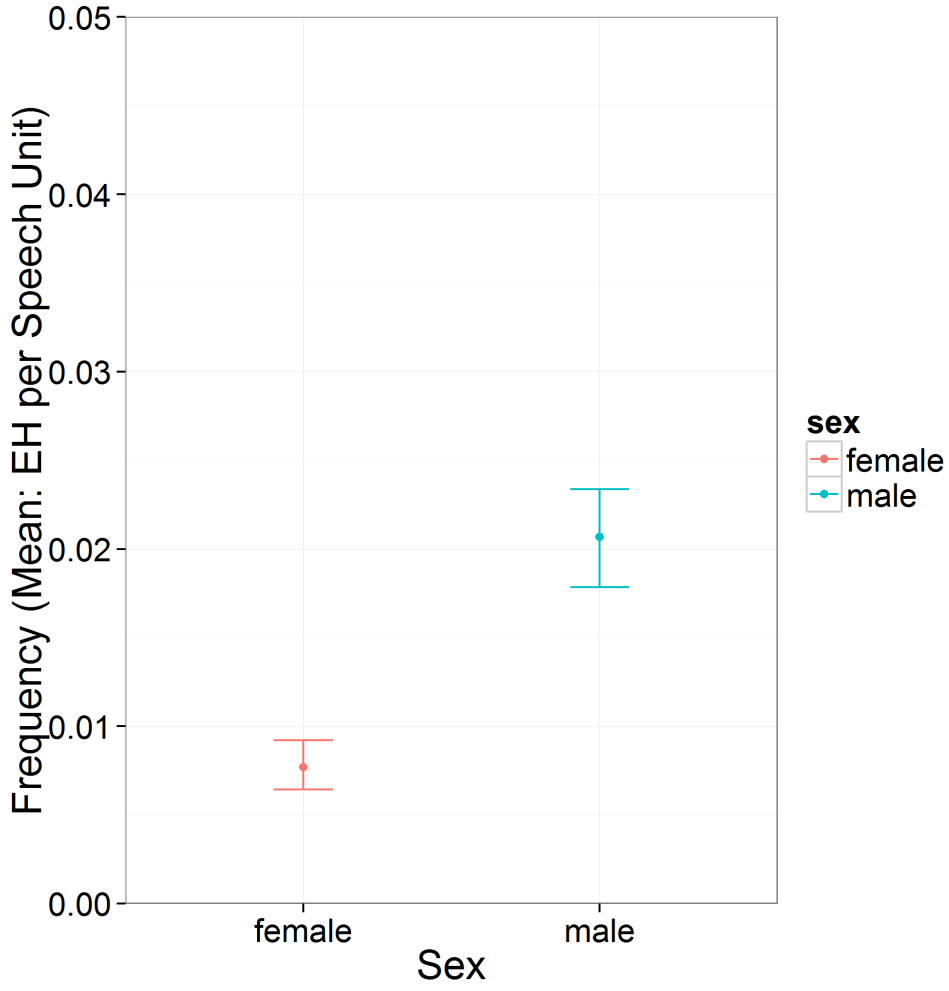
Figure: Linear versus nested variable structure

Table of Contents

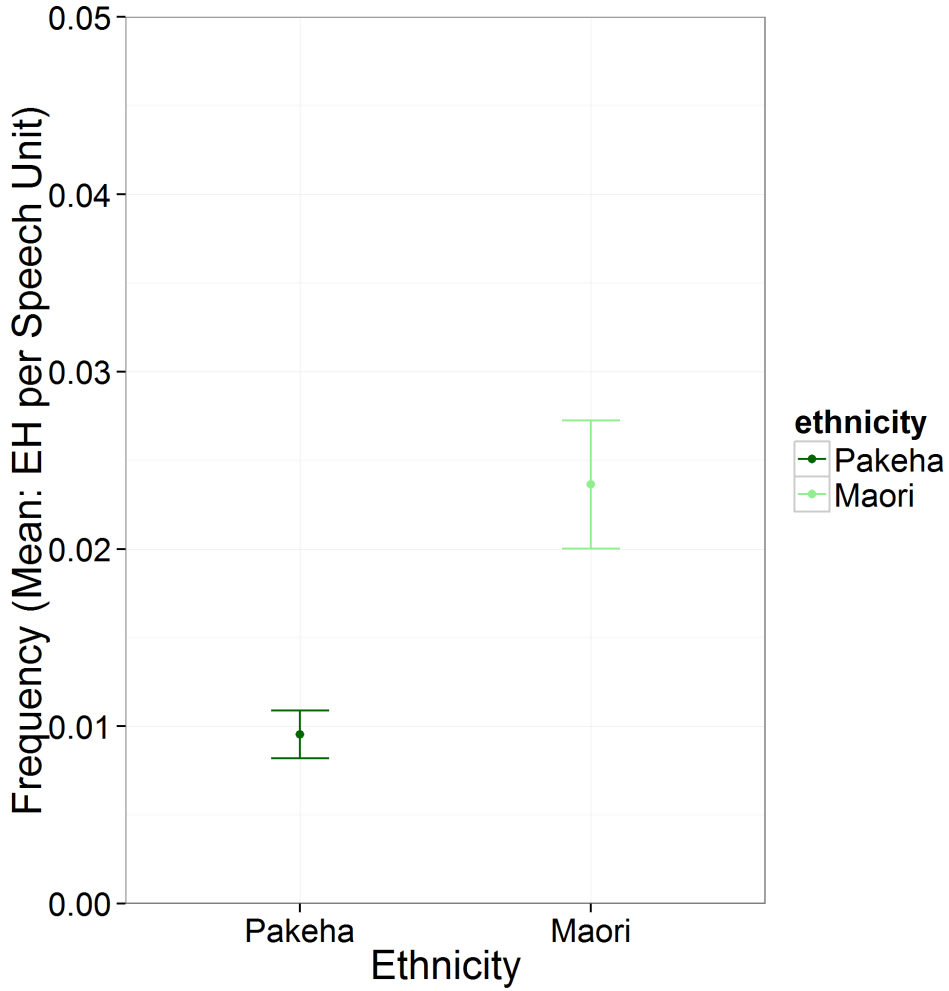
eh by age


0.05
0.04
0.03
0.02
0.01
0.00
Frequency (Mean: EH per Speech Unit)
young
old
Age
age
young
old

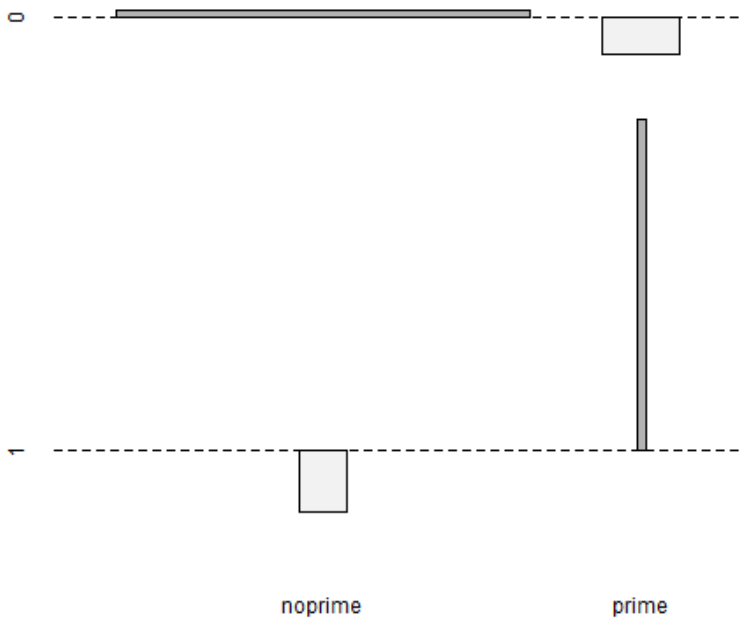
eh by gender


0.05
0.04
0.03
0.02
0.01
0.00
Frequency (Mean: EH per Speech Unit)
female
male
Sex
sex
female
male

eh by ethnicity



eh by priming



χ^2: 20.14, DF $= 1$, p-value $< .001$***

Factor	Bivariate Statistics	Binary Logistic Regression		Mixed-Effects Logistic Regression	
		Odds Ratio	Significance	Odds Ratio	Significance
Age	χ^2: 15.73, DF:1, p< .001***	0.39	p< .001***	0.33	p< .001***
Sex	χ^2: 81.03, DF:1, p< .001***	2.53	p< .001***	2.48	p< .001***
Ethnicity	χ^2: 70.09, DF:1, p< .001***	2.80	p< .001***	3.79	p< .001***
Priming	χ^2: 20.14, DF:1, p< .001***	1.58	p< .05*		n.s.
		Model statistics			
R2 (Nagelkerke)			0.054		0.086
AIC			3,515.87		3,240.68
BIC			3,336.54		3,281.48
Somer's Dxy			0.374		0.687
C			0.687		0.847

Table of Contents

Discussion

- Sociolinguistic interpretation
  - Strongest effect: ethnicity → identity marker among Maori
  - Males use *eh* more frequently than females → confirms Labov's "Gender Paradox", i.e. that women conform more closely than men to sociolinguistic norms that are overtly prescribed, but conform less than men when they are not. (Labov 2001: 292–293)
  - Young speakers use *eh* more than older speakers → typical pattern we observe in cases of ongoing change.
  - No significant correlation between *eh* use and priming
  - Effect of priming not relevant for ongoing change?

Discussion

- A statisticians interpretation
  - Bivariate statistics can be misleading: do not control for confounding factors.
  - Traditional statistical models tend to overestimate significance and the impact of extra-linguistic variables
  - May well be fatally flawed and may have led to flawed generalizations about the workings of language.

Discussion

- Remaining issues
  - Mixed-Effects Models also have shortcomings
    Extremely high β-error rate – that the model reports a factor as being not significant although it actually is in reality (cf. Johnson 2009: 368) (up to 90% if there are only few speakers, the effect size is very small, and there is a lot of within speaker variation).
  - Weighing options
    Is it better to say "We don't know because given the data we have we do not have enough evidence to say that factor X is significant" or to say "Factor X is significant" (although it is not)?

Table of Contents

Conclusion and Outlook

- This case study of *eh*
  - Distinct sociolinguistic profile
  - Example: improving our understanding of determinants of linguistic variability by using up-to-date statistics
  - Advanced statistical models outperform traditional fixed-effects models
  - More robust/reliable results
  - Allow including more predictors (converging socio- and psycholinguistics)
  - Impact on theorizing!

Gries, S. T. (2013). Sources of variability relevant to the cognitive sociolinguist, and corpus- as well as psycholinguistic methods and notions to handle them. *Journal of Pragmatics 52*, 5–16.

Johnson, D. E. (2009). Getting off the goldvarb standard: Introducing rbrul for mixed-effects variable rule analysis. *Language and linguistics compass 3*(1), 359–383.

Labov, W. (2001). *Principles of Linguistic Change: Social Factors*, Volume 2. Blackwell.

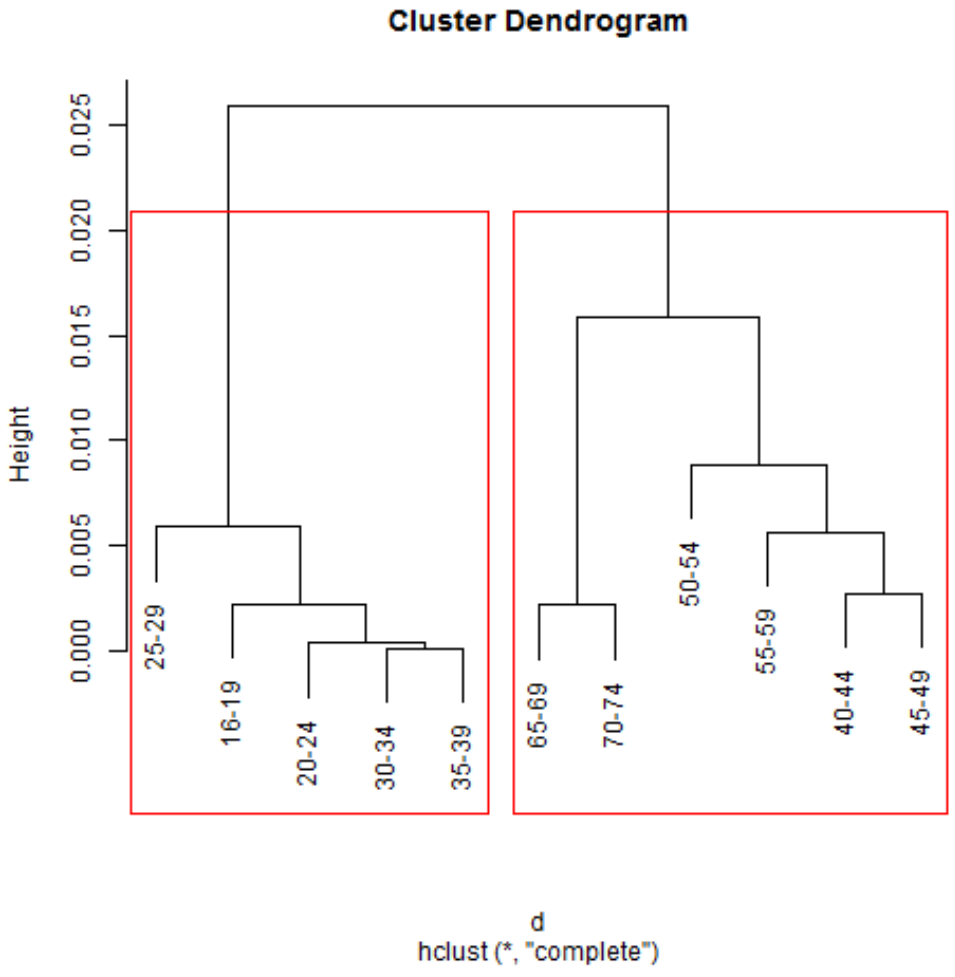
Meyerhoff, M. (1994). Sounds pretty ethnic, eh?: A pragmatic particle in new zealand english. *Language in Society 23*, 367–367.

Sankoff, D., S. A. Tagliamonte, and E. Smith (2005). Goldvarb x: A variable rule application for macintosh and windows.

Tulving, E. and D. L. Schacter (1990). Priming and human memory systems. *Science 247*(4940), 301–306.

Table of Contents

re-coding of age



	Estimate	VIF	OddsRatio	Std. Error	z value	$Pr(>\|z\|)$
(Intercept)	-4.97		0.01	0.1	-49.46	p< .001***
ageold	-0.95	1.07	0.39	0.16	-6.06	p< .001***
sexmale	0.93	1.02	2.53	0.12	7.99	p< .001***
ethnicityMaori	1.03	1.09	2.8	0.12	8.71	p< .001***
primingprime	0.46	1.03	1.58	0.21	2.16	p< .05*

Model statistics	Value
Null deviance	3515.87
Residual deviance	3336.54
R2 (Nagelkerke)	0.054
C	0.687
Somers' Dxy	0.374
AIC	3346.54
Prediction accuracy	98.73%

Model Likelihood Ratio Test	L.R.χ^2: 179.33	DF: 4	p-value: 0	sig: p< .001***

Base-line model: suf.eh∼1

Term Added	DF	Residual Deviance	p-value	Significance
age	1	17.91	0.00002	p< .001***
sex	1	82.63	0.00000	p< .001***
ethnicity	1	74.63	0.00000	p< .001***
priming	1	4.16	0.04149	p< .05*
age:sex	1	1.9	0.16806	n.s.
sex:ethnicity	1	0.55	0.45873	n.s.
age:ethnicity	1	0.99	0.32052	n.s.
age:sex:ethnicity	1	2.31	0.12820	n.s.

Full model: suf.eh∼age+sex+ethnicity+priming+age:sex+sex:ethnicity+ age:ethnicity+age:sex:ethnicity

Term Removed	DF	Residual Deviance	p-value	Significance
age:sex:ethnicity	1	-2.31	0.12820	n.s.
age:ethnicity	1	-0.99	0.32052	n.s.
sex:ethnicity	1	-0.55	0.45873	n.s.
age:sex	1	-1.90	0.16806	n.s.
priming	1	-4.16	0.04149	p< .05*
ethnicity	1	-74.63	0.00000	p< .001***
sex	1	-82.63	0.00000	p< .001***
age	1	-17.91	0.00002	p< .001***

	Group(s)	Variance	Std. Dev.	L.R. χ^2	DF	$Pr(>\|z\|)$
Random Effect(s)	file.speaker.id	1.04	1.02	237.2	1	p< .001***
Fixed Effect(s)	**Estimate**	**VIF**	**OddsRatio**	**Std. Error**	**z value**	$Pr(>\|z\|)$
(Intercept)	-5.48		0	0.18	-29.75	p< .001***
ageold	-1.1	1.11	0.33	0.28	-3.89	p< .001***
sexmale	0.91	1	2.48	0.21	4.27	p< .001***
ethnicityMaori	1.33	1.11	3.79	0.26	5.17	p< .001***
Model statistics						**Value**
Residual deviance						3,230.68
R2 (Nagelkerke)						0.086
C						0.843
Somers' Dxy						0.687
AIC						3,240.68
BIC						3,281.48
Prediction accuracy						98.73%
Model Likelihood Ratio Test			L.R. χ^2: 285.18	DF: 4	p-value: 0	sig: p< .001***

Base-line model: suf.eh∼(1|file.speaker.id)

Model	Term Added	Compared to...	DF	AIC	BIC	LogLikelihood	Residual Deviance	χ^2	χ^2DF	p-value	Significance
m1.glmer	age	m0.glmer	3	3278.87	3303.35	-1636.43	3272.87	5.80	1	0.01602	$p < .05$*
m2.glmer	sex	m1.glmer	4	3265.21	3297.85	-1628.61	3257.21	15.66	1	0.00008	$p < .001$***
m3.glmer	ethnicity	m2.glmer	5	3240.68	3281.48	-1615.34	3230.68	26.53	1	0.00000	$p < .001$***
m4.glmer	priming	m3.glmer	6	3242.65	3291.61	-1615.33	3230.65	0.03	1	0.86233	n.s.
m5.glmer	age:sex	m4.glmer	7	3243.63	3300.74	-1614.82	3229.63	1.02	1	0.31161	n.s.
m6.glmer	age:ethnicity	m5.glmer	8	3245.58	3310.85	-1614.79	3229.58	0.05	1	0.82867	n.s.
m7.glmer	sex:ethnicity	m6.glmer	9	3247.24	3320.67	-1614.62	3229.24	0.35	1	0.55646	n.s.
m8.glmer	age:sex:ethnicity	m7.glmer	10	3246.66	3328.25	-1613.33	3226.66	2.58	1	0.10809	n.s.

Full model: suf.eh∼age+sex+ethnicity+priming+age:sex+age:ethnicity+sex:ethnicity+age:sex:ethnicity+(1|file.speaker.id)

Model	Term Deleted	Compared to...	DF	AIC	BIC	LogLikelihood	Residual Deviance	χ^2	χ^2DF	p-value	Significance
m8.glmer	age:sex:ethnicity	m7.glmer	10	3246.66	3328.25	-1613.33	3226.66	2.58	1	0.10809	n.s.
m7.glmer	sex:ethnicity	m6.glmer	9	3247.24	3320.67	-1614.62	3229.24	0.35	1	0.55646	n.s.
m6.glmer	age:ethnicity	m5.glmer	8	3245.58	3310.85	-1614.79	3229.58	0.05	1	0.82867	n.s.
m5.glmer	age:sex	m4.glmer	7	3243.63	3300.74	-1614.82	3229.63	1.02	1	0.31161	n.s.
m4.glmer	priming	m3.glmer	6	3242.65	3291.61	-1615.33	3230.65	0.03	1	0.86233	n.s.
m3.glmer	ethnicity	m2.glmer	5	3240.68	3281.48	-1615.34	3230.68	26.53	1	0.00000	$p < .001$***
m2.glmer	sex	m1.glmer	4	3265.21	3297.85	-1628.61	3257.21	15.66	1	0.00008	$p < .001$***
m1.glmer	age	m0.glmer	3	3278.87	3303.35	-1636.43	3272.87	5.80	1	0.01602	$p < .05$*

Thank you very much!

Why We Need Better Statistics –
A Case Study on *eh* in New Zealand English

www.martinschweinberger.de
University of Hamburg