Exploring the Coding Dynamics: Evaluating the Association between Race and Blood Pressure

Victoria Grase, BSc[1], Anarina Murillo, PhD[1], Stavroula A. Chrysanthopoulou, PhD[1]
[1]Brown University School of Public Health Biostatistics Department, RI,

Overview

The aim of this study is to compare four distinct race coding methods to assess the association between systolic blood pressure (BP) and blood lead levels (BLL) adjusted for race.

Background

- Higher BP in non-Hispanic (NH) black (56%) than NH white (48%), NH Asian (46%), Hispanic (39%) adults.
- BP medication led to better BP control in NH white (32%), NH black (25%), Asian (19%), or Hispanic (25%) adults.
- Experiencing discrimination is related to higher BP according to the National Heart, Lung, and Blood Institute.
- Typically, racial differences are examined using dummy coding however, other race coding methods allow for evaluating alternative hypotheses.

Study Design

- Multiple regression models were applied to predict BP based on BLL adjusted for race using NHANES data.
- Race Groups: 1=Mexican American, 2=Other Hispanic, 3=NH White, 4=NH Black, 5=NH Multiracial
- Race coding methods:

Simple Coding:
Represent category membership by incorporating a set of variables a set of variables coded (-1,0,1) allows you to compare means at each level of a variable to the mean of another.

SIMPLE CODING	NEW VARIABLE 1 (SX1)	NEW VARIABLE 2 (SX2)	NEW VARIABLE 3 (SX3)	NEW VARIABLE 4 (SX4)
Group 1	-1 / k	-1 / k	-1 / k	-1 / k
Group 2	(k-1) / k	-1 / k	-1 / k	-1 / k
Group 3	-1 / k	(k-1) / k	-1 / k	-1 / k
Group 4	-1 / k	-1 / k	(k-1) / k	-1 / k
Group 5	-1 / k	-1 / k	-1 / k	(k-1) / k

Difference (Forward or Backward) Coding:
Compares the mean of each level of the variable to the mean of the previous level. More useful for ordinal variables, does not make sense for race.

FORWARD DIFFERENCE CODING	NEW VARIABLE 1 (FX1)	NEW VARIABLE 2 (FX2)	NEW VARIABLE 3 (FX3)	NEW VARIABLE 4 (FX4)
	Group 1 v. Group 2	Group 2 v. Group 3	Group 3 v. Group 4	Group 4 v. Group 5
Group 1	(k-1)/k	(k-2)/k	(k-3)/k	(k-4)/k
Group 2	-1 / k	(k-2)/k	(k-3)/k	(k-4)/k
Group 3	-1 / k	-2 / k	(k-3)/k	(k-4)/k
Group 4	-1 / k	-2 / k	-3 / k	(k-4)/k
Group 5	-1 / k	-2 / k	-3/ k	-4 / k

BACKWARD DIFFERENCE CODING	NEW VARIABLE 1 (BX1)	NEW VARIABLE 2 (BX2)	NEW VARIABLE 3 (BX3)	NEW VARIABLE 4 (BX4)
	Group 1 v. Group 2	Group 2 v. Group 3	Group 3 v. Group 4	Group 4 v. Group 5
Group 1	-(k-1)/k	-(k-2)/k	-(k-3)/k	-(k-4)/k
Group 2	1 / k	-(k-2)/k	-(k-3)/k	-(k-4)/k
Group 3	1 / k	2 / k	-(k-3)/k	-(k-4)/k
Group 4	1 / k	2 / k	3 / k	-(k-4)/k
Group 5	1 / k	2 / k	3/ k	4 / k

Deviation Coding:
Compares mean of the dependent variable for each level of the coded variable to the group of other levels of that variable.

DEVIATION CODING	NEW VARIABLE 1 (DX1)	NEW VARIABLE 2 (DX2)	NEW VARIABLE 3 (DX3)	NEW VARIABLE 4 (DX4)
	Group 2 v. Mean	Group 3 v. Mean	Group 4 v. Mean	Group 5 v. Mean
Group 1	-1	-1	-1	-1
Group 2	1	0	0	0
Group 3	0	1	0	0
Group 4	0	0	1	0
Group 5	0	0	0	1

Results

- A sample of 24,429 participants were analyzed from NHANES data (1999-2017) aged > 20 years.
- Individuals with missing BP and insurance were excluded.
- Demographic, insurance, smoking, alcohol, BMI, hypertension, blood lead levels, and BP were included.

CODING METHODS	BLL	MEXICAN AMERICAN	OTHER HISPANIC	NON-HISPANIC WHITE	NON-HISPANIC BLACK	NON-HISPANIC MULTIRACIAL
Simple Coding	1.51 (1.37,1.63) ***	121.13 (120.46,121.79) ***	-0.46 (-1.46,0.54)	0.74 (0.10,1.39) *	4.69 (3.93,5.46) ***	-1.65 (-2.72,-0.58) **
Male	0.89 (0.75,1.02) ***	122.78 (122.01,123.54) ***	-0.18 (-1.47,1.12)	0.44 (-0.36,1.25)	3.98 (3.02,4.94) ***	-1.86 (-3.19,-0.53) **
Female	3.67 (3.38,3.96) ***	114.74 (113.75,115.73) ***	-0.45 (-1.99,1.10)	1.03 (0.02,2.06) *	5.14 (3.94,6.34) ***	-1.32 (-3.05,0.41)
Forward Coding	1.51 (1.37,1.63) ***	121.79 (121.33,122.25) ***	0.46 (-0.55,1.46)	-1.20 (-2.10,-0.30) **	-3.95 (-4.56,-3.34) ***	6.34 (5.29,7.40) ***
Male	0.89 (0.75,1.02) ***	123.25 (122.77,123.73) ***	0.18 (-1.12,1.47)	-0.62 (-1.79,0.55)	-3.54 (-4.33,-2.75) ***	5.84 (4.52,7.16) ***
Female	3.67 (3.38,3.96) ***	115.62 (115.01,116.24) ***	0.45 (1.10,1.99)	-1.49 (-2.84,-0.13) *	-4.10 (-5.04,-3.16) ***	6.46 (4.77,8.15) ***
Backward Coding	1.51 (1.37,1.63) ***	121.79 (121.33,122.25) ***	-0.46 (-1.46,0.55)	1.20 (0.30,2.10) **	3.95 (3.34,4.56) ***	-6.34 (-7.40,-5.29) ***
Male	0.89 (0.75,1.02) ***	123.25 (122.77,123.73) ***	-0.18 (-1.47,1.12)	0.62 (-0.55,1.79)	3.54 (2.75,4.33) ***	-5.84 (-7.16,-4.52) ***
Female	3.67 (3.38,3.96) ***	115.62 (115.01,116.24) ***	-0.45 (-1.99,1.10)	1.49 (0.13,2.84) *	4.10 (3.16,5.04) ***	-6.46 (-8.15,-4.77) ***
Deviation Coding	1.51 (1.37,1.63) ***	121.79 (121.33,122.25) ***	-1.12 (-1.84,-0.41) **	0.08 (-0.31,0.47)	4.03 (3.52,4.53) ***	-2.32 (-3.09,-1.54) ***
Male	0.89 (0.75,1.02) ***	123.25 (122.77,123.73) ***	-0.65 (-1.58,0.28)	-0.04 (-0.53,0.46)	3.51 (2.86,4.15) ***	-2.34 (-3.30,-1.38) ***
Female	3.67 (3.38,3.96) ***	115.62 (115.01,116.24) ***	-1.33 (-2.41,-0.25) *	0.16 (-0.45,0.76)	4.26 (3.48,5.04) ***	-2.20 (-3.44,-0.96) ***

The overall difference estimate between male and female is -3.03.

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Conclusion

- NH Black individuals consistently show the highest systolic blood pressures across different coding schemes, irrespective of gender.
- Across all coding scheme methods and models male continually exhibited higher systolic blood pressure than females.
- The best model, which included education and age, demonstrated superior predictive power in estimating blood pressure (not shown).
- There was no statistical significance between the interaction term for BLL and race in nearly all coding schemes or groups (not shown).
- Our investigation highlights consistent significance across all race coding methods.
- Careful consideration of coding methods is essential to align statistical analysis with the hypothesis under investigation.
- Scrutinizing race coding's impact on estimations, this research illuminates nuanced dynamics in health associations, aiding more effective interventions for health equity.
- Future work
  - Researchers should consider applying coding methods to other categorical variables.
  - Add in other variables to best model such as BMI, income, smoking, and/or alcohol.

Key References

1. *Goodman, M. S. (2022, June). Methodological issues and strategies for using and measuring race in regression analyses. In Biostatistics and Epidemiology Virtual Symposium 2022 (No. 08). Stata Users Group.*
2. *Huang Z. (2022). Association Between Blood Lead Level With High Blood Pressure in US (NHANES 1999-2018). Frontiers in public health, 10, 836357. https://doi.org/10.3389/fpubh.2022.836357*
3. *Alkharusi, H. (2012). Categorical variables in regression analysis: A comparison of dummy and effect coding. International Journal of Education, 4(2), 202.*