

Title: Modeling the Structure of Multi-racial Identity: A Latent Class Approach  
Ross Macmillan and Carolyn Liebler  
University of Minnesota

The new data on race and Hispanic origin provide a vast array of information about the racial and ethnic composition of the United States. Yet, social scientists, demographers, and policy makers are now faced with the serious issue of using these data appropriately. While data collection is well-advanced, theoretical and methodological developments are only in the beginning stages. To contribute to this area, this paper articulates the utility of latent class analysis for modeling such phenomena as a basis for theoretical advance. Specifically, we examine the latent structure of race and Hispanic origin responses in Census 2000 to provide some empirical basis for further theorizing of America's current race/ethnic situation. For example, race theorists have repeatedly hypothesized that America's most powerful racial divide today is between Blacks and non-Blacks, rather than between Whites and non-Whites. Our analyses will provide a new testing ground for this and related ideas because we take into account the full array of multiple race and Hispanic origin responses.

The Census 2000 data provide six dichotomous variables indicating whether each individual reported their race as White, Black/African American, American Indian/Alaska Native, Asian, Pacific Islander, or Other. We will also create a dichotomous variable indicating that Hispanic origin was reported.<sup>1</sup> Taking all seven of the dichotomous indicators together gives  $2^7$  or 128 potential combinations of responses. The frequencies of each of these responses in the 2000 5% PUMS data are shown in Table 1, below. Moreover, the 2000 Census provides one of the few data sets in existence that has a large enough sample to systematically analyze these many combinations. To this end, our analyses proceed in two parts.

First, we will conduct latent class analysis on the 128 categories of responses given in the Census 2000 5% PUMS data for the United States as a whole. This allows us to systematically analyze the underlying or latent structure of both mono- and multi-racial identity. Such analyses generate estimates of conditional probabilities of multiple observed racial indicators given latent class. These serve as central descriptors of multi-racial identity as they show the probabilistic clustering of racial indicators within empirically homogenous groupings. These analyses also provide latent class probabilities that indicate the expected proportions of a population that are characterized by particular clusters of observed racial identities. Ultimately, such models reveal the underlying or latent structure of racial identity and provide a basis for theorizing the social structure of racial identity in the United States.

Second, we will divide the country into regions and conduct the simultaneous group-specific latent class analyses for the different regions. Frequencies of race responses by region are also presented in Table 1. We conduct formal comparisons of homogeneous models in which the latent structure of multi-racial identity is constrained to be equal, partially homogeneous models in which particular aspects (i.e., conditional probabilities of particular racial identifiers) are constrained to be equal across regions, and fully heterogeneous models in which latent structure is free to vary. These analyses test whether the classes differ significantly between regions and

---

<sup>1</sup> For people whose responses to the race question and the Hispanic origin question were allocated, we will focus only on the fact that they were allocated and not on the particular allocated race/origin.

allow assessment of how various regional factors, including immigration, interracial unions, and selective race reporting, structure the nature of multi-racial identity. We first focus on Census regions, but will move to more theoretically derived regional divisions (if time permits).

This paper also includes a technical discussion of latent class analysis and various issues (i.e., identification, implications of sparse data, model specification, model selection, and modeling strategy) that researchers are likely to encounter in other applications.

We conclude the paper by elaborating the theoretical implications of our research and discussing further applications of the methodology for substantive questions in the field.

Table 1:  
**Frequency of each Race/Hispanic Origin Response, by Region**  
 Data: Census 2000 5% PUMS

		Total in US	northeast	midwest	south	west
0	race & Hisp allocated	6,172,871	1,224,525	1,075,334	1,936,869	1,936,143
1	W	195655615	39,335,142	52,048,105	66,727,771	37,544,597
2	B	32981565	5,616,127	6,213,834	18,288,633	2,862,971
3	W, B	716,077	136,903	192,607	217,124	169,443
4	AI	2,129,992	131,560	373,062	650,811	974,559
5	W, AI	1,146,729	106,849	250,860	422,464	366,556
6	B, AI	190,747	44,155	34,420	69,201	42,971
7	W, B, AI	98,015	19,155	22,033	26,225	30,602
8	A	9,782,495	2,035,861	1,130,035	1,847,206	4,769,393
9	W, A	807,891	106,504	110,775	183,883	406,729
10	B, A	99,051	19,156	13,340	32,567	33,988
11	W, B, A	18,623	3,825	2,414	5,303	7,081
12	AI, A	43,218	10,386	5,357	8,626	18,849
13	W, AI, A	17,041	1,464	1,859	4,235	9,483
14	B, AI, A	5,099	906	759	1,631	1,803
15	W, B, AI, A	8,495	1,407	1,668	2,338	3,082
16	PI,	325,182	17,506	20,477	40,306	246,893
17	W, PI,	101,337	9,309	10,983	16,453	64,592
18	B PI,	29,926	8,681	3,065	9,928	8,252
19	W, B PI,	2,526	606	212	831	877
20	AI, PI,	6,157	614	709	1,377	3,457
21	W, AI, PI,	4,193	435	616	878	2,264
22	B, AI, PI,	708	181	226	69	232
23	W, B, AI, PI,	874	112	216	292	254
24	A, PI	124,411	9,642	10,054	13,901	90,814
25	W, A, PI,	79,423	3,655	5,024	7,225	63,519
26	B, A, PI,	4,347	486	579	1,552	1,730
27	W, B, A, PI,	1,860	344	210	546	760
28	AI, A, PI,	2,328	334	271	689	1,034
29	W, AI, A, PI,	4,242	298	474	304	3,166
30	B, AI, A, PI,	268	36	79	50	103
31	W, B, AI, A, PI,	3,287	466	584	1,102	1,135
32	OT	3,799,411	756,712	375,214	1,000,460	1,667,025
33	W, OT	1,155,137	256,026	158,731	273,815	466,565
34	B, OT	321,194	134,029	30,980	117,374	38,811
35	W, B, OT	35,491	8,667	6,853	11,422	8,549
36	AI, OT	44,030	9,042	4,001	10,064	20,923
37	W, AI, OT	22,904	2,388	3,293	6,186	11,037
38	B, AI, OT	6,649	1,592	689	1,403	2,965
39	W, B, AI, OT	2,796	382	349	1,243	822
40	A, OT	209,126	55,673	20,228	45,310	87,915
41	W, A, OT	30,313	3,602	2,369	5,470	18,872
42	B, A, OT	6,555	2,116	802	2,040	1,597
43	W, B, A, OT	1,095	324	162	267	342
44	AI, A, OT	1,402	226	179	470	527
45	W, AI, A, OT	792	74	180	373	165
46	B, AI, A, OT	299	94	36	37	132
47	W, B, AI, A, OT	521	73	73	188	187
48	PI, OT	21,347	5,050	2,010	3,982	10,305
49	W, PI, OT	6,199	366	379	992	4,462
50	B PI, OT	1,133	213	134	451	335
51	W, B PI, OT	189	0	25	65	99
52,	AI, PI, OT	151	52	17	56	26
53	W, AI, PI, OT	281	11	26	117	127
54	B, AI, PI, OT	196	0	14	88	94
55	W, B, AI, PI, OT	0	0	0	0	0
56	A, PI, OT	2,947	366	238	600	1,743
57	W, A, PI, OT	8,245	256	287	405	7,297
58	B, A, PI, OT	457	137	0	218	102
59	W, B, A, PI, OT	84	0	67	11	6
60	AI, A, PI, OT	107	0	66	35	6
61	W, AI, A, PI, OT	378	58	18	47	255
62	B, AI, A, PI, OT	65	0	0	37	28

63	W, B, AI, A, PI, OT	249	60	0	157	32
64	H (race allocated)	2,962,968	471,910	271,496	716,134	1,503,428
65	W, H	10811139	1,306,177	905,492	4,557,086	4,042,384
66	B, H	386,402	168,874	35,895	115,462	66,171
67	W, B, H	58,186	21,989	6,647	14,718	14,832
68	AI, H	222,248	22,596	22,264	49,172	128,216
69	W, AI, H	85,938	6,127	10,723	19,065	50,023
70	B, AI, H	10,245	3,637	1,317	2,275	3,016
71	W, B, AI, H	13,555	3,819	2,067	2,956	4,713
72	A, H	66,077	9,110	5,264	11,777	39,926
73	W, A, H	39,496	2,737	3,540	7,918	25,301
74	B, A, H	3,483	809	512	768	1,394
75	W, B, A, H	2,528	474	307	1,139	608
76	AI, A, H	4,647	562	353	884	2,848
77	W, AI, A, H	3,567	376	469	602	2,120
78	B, AI, A, H	600	57	74	331	138
79	W, B, AI, A, H	706	169	0	250	287
80	PI, H	18,420	1,503	1,759	3,516	11,642
81	W, PI, H	7,832	828	1,301	886	4,817
82	B, PI, H	1,427	274	131	712	310
83	W, B PI, H	251	11	66	23	151
84	AI, PI, H	904	200	275	107	322
85	W, AI, PI, H	625	54	58	78	435
86	B, AI, PI, H	497	94	51	0	352
87	W, B, AI, PI, H	208	84	82	42	0
88	A, PI, H	5,703	188	151	545	4,819
89	W, A, PI, H	8,359	325	416	788	6,830
90	B, A, PI, H	431	58	13	163	197
91	W, B, A, PI, H	322	58	54	50	160
92	AI, A, PI, H	657	33	69	149	406
93	W, AI, A, PI, H	1,201	55	168	114	864
94	B, AI, A, PI, H	232	33	175	11	13
95	W, B, AI, A, PI, H	546	118	16	108	304
96	OT, H	9,206,353	1,282,325	831,931	2,386,483	4,705,614
97	W, OT, H	938,170	135,613	96,787	267,771	437,999
98	B, OT, H	117,887	39,619	12,003	34,467	31,798
99	W, B, OT, H	14,627	4,948	2,039	2,805	4,835
100	AI, OT, H	56,499	7,486	4,747	11,793	32,473
101	W, AI OT, H	15,423	1,590	1,731	3,602	8,500
102	B, AI OT, H	1,714	457	230	426	601
103	W, B, AI OT, H	2,280	484	273	812	711
104	A, OT, H	56,752	7,902	4,806	10,350	33,694
105	W, A, OT, H	15,000	1,537	1,391	2,034	10,038
106	B, A, OT, H	1,681	417	275	387	602
107	W, B, A, OT, H	511	138	52	271	50
108	AI, A, OT, H	833	118	72	311	332
109	W, AI, A, OT, H	529	89	165	100	175
110	B, AI, A, OT, H	56	0	0	37	19
111	W, B, AI, A, OT, H	203	80	40	63	20
112	PI, OT, H	19,424	5,588	2,529	3,251	8,056
113	W, PI, OT, H	3,421	393	524	663	1,841
114	B PI, OT, H	589	113	129	149	198
115	W, B PI, OT, H	167	41	43	62	21
116	, AI, PI, OT, H	163	13	28	37	85
117	W, AI, PI, OT, H	237	51	16	36	134
118	B, AI, PI, OT, H	86	45	0	0	41
119	W, B, AI, PI, OT, H	0	0	0	0	0
120	A, PI, OT, H	2,339	176	309	486	1,368
121	W, A, PI, OT, H	2,440	71	145	268	1,956
122	B, A, PI, OT, H	107	0	41	66	0
123	W, B, A, PI, OT, H	137	30	59	0	48
124	AI, A, PI, OT, H	70	28	0	42	0
125	W, AI, A, PI, OT, H	371	23	0	103	245
126	B, AI, A, PI, OT, H	45	45	0	0	0
127	W, B, AI, A, PI, OT, H	56	0	0	44	12

**Legend:** W=White; B=Black or African American; AI=American Indian or Alaska Native; A=Asian; PI=Native Hawaiian or Other Pacific Islander; OT=Some other race; H=of Hispanic origin