Name Method: tables for the manuscript

Cal Chengqi Fang

2024-06-27

In this document, I made some tables for our manuscript to illustrate our analysis process and results.

Table 1. Illustrative surnames and corresponding conditional probabilities

Country/Area of origin	Illustrative surnames[*]								
	HUANG	CHANG	KIM	SUZUKI	NGUYEN	INTHAVONG	PENAFLOR	SAMEER	PATAN
Afghanistan	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	17.74
Bangladesh, India, Pakistan, Sikkim, British India	0.09	0.24	0.31	0.00	0.02	0.00	0.01	52.13	33.87
Bhutan	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Brunei, Indonesia	0.13	0.23	0.02	0.03	0.14	0.00	0.01	0.00	0.00
Cambodia	0.07	0.37	0.74	0.03	0.32	0.00	0.01	0.00	0.00
Japan, Southern Ryukyu Islands	0.19	0.62	0.81	95.73	0.07	0.00	0.01	0.00	0.00
Korea, Democratic People's Republic of Korea, Republic of Korea	0.13	21.64	94.74	0.32	0.07	0.00	0.01	0.00	0.00
Laos	0.08	3.02	0.01	0.00	0.31	90.75	0.01	0.00	0.00
Malaysia	0.12	0.88	0.05	0.00	0.17	0.00	0.01	0.00	0.00
Maldives	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00
Mongolia	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Myanmar	0.14	0.50	0.15	0.00	0.01	0.00	0.01	0.00	0.00
Nepal	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Peoples Republic of China, Hong Kong, Macau (Macao), Taiwan	96.28	62.10	0.34	0.60	0.25	0.00	0.01	0.00	0.00
Philippines	0.32	0.56	0.10	0.33	0.41	0.91	69.13	0.00	0.00
Singapore	0.22	0.35	0.02	0.04	0.01	0.00	0.01	0.00	0.00
Sri Lanka	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Thailand	0.06	2.21	0.11	0.09	0.20	7.88	0.01	0.00	6.45
Vietnam, Democratic Republic of Vietnam, Republic of Vietnam	1.31	1.52	0.59	0.03	95.14	0.45	0.01	0.00	0.00
All the other countries and regions	0.86	5.75	2.02	2.79	2.88	0.00	30.65	47.87	41.94

[*] The probabilities in each column do not add up to 1 because of rounding error.

Table 2. Predictive surnames and related record numbers

Country/Area of origin	Number of predictive surnames[*]	Number of US-born records with corresponding predictive surnames	Number of foreign-born records with corresponding predictive surnames
Afghanistan	215	12,946	18,672
Bangladesh, India, Pakistan, Sikkim, British India	6,738	1,281,803	2,506,632
Bhutan	58	4,628	22,019
Brunei, Indonesia	368	14,793	41,692
Cambodia	537	96,110	92,847
Japan, Southern Ryukyu Islands	2,995	523,987	758,857
Korea, Democratic People's Republic of Korea, Republic of Korea	277	998,728	1,289,503
Laos	574	158,842	127,920
Malaysia	119	7,105	19,192
Maldives	0	0	0
Mongolia	48	1,060	3,236
Myanmar	187	19,991	71,161
Nepal	138	12,000	63,997
Peoples Republic of China, Hong Kong, Macau (Macao), Taiwan	572	1,078,677	2,813,805
Philippines	7,794	940,978	923,734
Singapore	11	466	1,852
Sri Lanka	115	10,020	15,979
Thailand	41	1,088	2,524
Vietnam, Democratic Republic of Vietnam, Republic of Vietnam	257	594,284	1,221,744

[*] The numbers do not add up to 22,621 because there are surnames that are equally predictive of more than one country/area.

Table 3. Validation result with ACS data

	Number of surnames present in both lists[*]	PPV (%)	Sensitivity (%)	Specificity (%)
Whole List	3170	77.41	87.02	97.99
Afghanistan	23	4.35	NA	NA
Bangladesh, India, Pakistan, Sikkim, British India	908	81.83	NA	NA
Bhutan	15	93.33	NA	NA
Brunei, Indonesia	21	61.90	NA	NA
Cambodia	96	69.79	NA	NA
Japan, Southern Ryukyu Islands	467	88.44	NA	NA
Korea, Democratic People's Republic of Korea, Republic of Korea	97	72.16	NA	NA
Laos	34	85.29	NA	NA
Malaysia	22	63.64	NA	NA
Maldives	0	NA	NA	NA
Mongolia	2	50.00	NA	NA
Myanmar	37	67.57	NA	NA
Nepal	18	94.44	NA	NA
Peoples Republic of China, Hong Kong, Macau (Macao), Taiwan	203	80.79	NA	NA
Philippines	1151	70.55	NA	NA
Singapore	3	66.67	NA	NA
Sri Lanka	8	62.50	NA	NA
Thailand	1	0.00	NA	NA
Vietnam, Democratic Republic of Vietnam, Republic of Vietnam	72	93.06	NA	NA

[*] The numbers of surnames of countries/areas do not add up to 3,370 because there are surnames that are equally predictive of more than one country/area.

Table 4. Validation result of cross validation

	PPV (%)	Sensitivity (%)	Specificity (%)
Afghanistan	75.95	51.44	99.96
Bangladesh, India, Pakistan, Sikkim, British India	87.40	98.84	95.94
Bhutan	62.42	47.71	99.92
Brunei, Indonesia	82.72	68.26	99.93
Cambodia	75.96	62.79	99.78
Japan, Southern Ryukyu Islands	94.56	96.65	99.55
Korea, Democratic People's Republic of Korea, Republic of Korea	80.52	93.51	97.17
Laos	71.86	71.86	99.63
Malaysia	73.01	22.44	99.95
Maldives	33.33	1.16	100.00
Mongolia	97.00	93.01	100.00
Myanmar	72.89	66.73	99.81
Nepal	72.61	72.41	99.82
Peoples Republic of China, Hong Kong, Macau (Macao), Taiwan	87.96	93.62	95.39
Philippines	86.74	95.03	98.67
Singapore	65.65	5.18	99.99
Sri Lanka	79.63	75.10	99.97
Thailand	86.64	2.84	100.00
Vietnam, Democratic Republic of Vietnam, Republic of Vietnam	91.49	92.67	98.82

Sample Table. Numbers of foreign-born records from corresponding country/area - Leftout

Country/Area of birth	Number of foreign-born records
Afghanistan	32,139
Bangladesh, India, Pakistan, Sikkim, British India	2,273,636
Bhutan	33,841
Brunei, Indonesia	58,115
Cambodia	136,556
Japan, Southern Ryukyu Islands	749,436
Korea, Democratic People's Republic of Korea, Republic of Korea	1,164,951
Laos	132,118
Malaysia	105,638
Maldives	150
Mongolia	3,502
Myanmar	98,144
Nepal	73,658
Peoples Republic of China, Hong Kong, Macau (Macao), Taiwan	2,768,925
Philippines	890,207
Singapore	46,230
Sri Lanka	18,651
Thailand	85,474
Vietnam, Democratic Republic of Vietnam, Republic of Vietnam	1,263,853
All the other countries and regions	668,407