

## Homework 3

Due 8am Wednesday Sept 28, 2016

Econ B2000, MA Econometrics

Kevin R Foster, CCNY

Each student should submit a separate assignment, even if it is an identical computer file to the rest of your study group. When submitting assignments, please include your name and the assignment number as part of the filename. Please write the names of your study group members at the beginning of your homework.

1. What are the names of the people currently in your study group?
2. Consider the average incomes of a particular subset (you choose) of people living in the city versus the rest of the state of New York. What is the difference in averages? What is the standard error of this difference? What is a 95% confidence interval? Are they statistically significantly different?
3. Use the PUMS data provided to find out something interesting about incomes, housing costs, and borough.
4. Try the machine learning K-nearest-neighbor algorithm on the PUMS data to get another view of the "interesting pattern" from above. (As usual, step one is to replicate my code, then gradually morph it into your own.) How good of a classification can you get?
5. (extra) The PUMS data includes a code, ANCESTR1D, for each person to give detail about their ancestry. The complete codes are below. You might figure out some recoding (note that most of the codes less than 2100 are European, then rest of 2000s are Latin America, 3000 are Caribbean, 4000 MENA, 5000 African, 6000 Indian subcontinent, 7000 Asia, 8000 Pacific, 9000 other). You might make judgements about how to combine groups. Look at average incomes within some of these ancestry groups, are they statistically significantly different? Explain.

ANCESTR1D	0100 "Walloon"	0213 "Dutch-Irish-Scotch"	0260 "French (1980)"	German (1990-2000, ACS, PRCS)"
0010 "Alsatian"	0110 "British"	0220 "English"	0261 "French (1990-2000, ACS, PRCS)"	0323 "East German (1990-2000)"
0020 "Andorran"	0120 "British Isles"	0221 "English-French-German"	0262 "Occitan (1990-2000)"	0324 "West German (1990-2000)"
0030 "Austrian"	0130 "Channel Islander"	0222 "English-French-Irish"	0270 "Lorrainian"	0325 "German-French-Irish"
0040 "Tirolean"	0140 "Gibraltan"	0223 "English-German-Irish"	0280 "Breton"	0326 "German-Irish-Italian"
0051 "Basque (1980)"	0150 "Cornish"	0224 "English-German-Swedish"	0290 "Frisian"	0327 "German-Irish-Scotch"
0052 "Spanish Basque (1980)"	0160 "Corsican"	0225 "English-Irish-Scotch"	0300 "Friulian"	0328 "German-Irish-Swedish"
0053 "Basque (1990-2000, ACS)"	0170 "Cypriot"	0226 "English-Dutch"	0321 "German (1990-2000, ACS, PRCS)"	0329 "Germanic"
0054 "Spanish Basque (1990-2000, 2001-2004 ACS)"	0180 "Greek Cypriote"	0227 "English-Faeroe Islander"	0322 "Pennsylvania"	0330 "Bavarian"
0060 "French Basque"	0190 "Turkish Cypriote"	0228 "Finnish-Karelian"		0340 "Berliner"
0080 "Belgian"	0200 "Danish"			
0090 "Flemish"	0210 "Dutch"			
	0211 "Dutch-French-Irish"			
	0212 "Dutch-German-Irish"			

0350 "Hamburger"	0760 "Liechtensteiner"	1111 "Czech"	(2000, ACS, PRCS)	2001 "Spaniard
0360 "Hanoverian"		1120 "Bohemian"	"	(1990-2000, ACS, PRCS)"
0370 "Hessian"	0770 "Luxemburger"	(1980)"	1523	2002 "Castillan
0380 "Lubecker"	0780 "Maltese"	1121 "Bohemian"	"Montenegrin	(1990-2000, 2012
0390 "Pomeranian"	0790 "Manx"	(1990-2000, ACS, PRCS)"	ACS)"	2003 "Velencian
(1980)"		1122 "Moravian"	1530 "Slovak"	(1990-2000)"
0391 "Pomeranian	0800 "Monegasque"	(1990-2000)"	1540 "Slovene"	2010 "Andalusian
(1990-2000)"	0810 "Northern	1150 "Estonian"	1550 "Sorb/Wend"	(1990-2000)"
0392 "Silesian"	Irlander"	1160 "Livonian"	1560 "Soviet	2020 "Asturian
(1990-2000)"		1170 "Finno	Turkic (1990-	(1990-2000)"
0400 "Prussian"	0820 "Norwegian"	Ugrian (1990-	2000)"	2040 "Catalonian"
0410 "Saxon"	0840 "Portuguese"	2000)"	1570 "Bashkir"	2050 "Balearic
0420 "Sudetenlander"	0850 "Azorean"	1171 "Udmert"	1580 "Chevash"	Islander (1980)"
0430 "Westphalian"	0860 "Madeiran"	1180 "Mordovian"	1590 "Gagauz	2051 "Balearic
0460 "Greek"	0870 "Scotch Irish"	1190 "Voytak"	(1990-2000)"	Islander (1990-
0470 "Cretan"	0880 "Scottish"	1200 "Georgian"	1600 "Meskhetian	2000)"
0480 "Cycladic	0890 "Swedish"	1220 "Germans	(1990-2000)"	2052 "Canary
Islander,	0900 "Aland	from Russia"	1630 "Yakut"	Islander (1990-
Dodecanese	0910 "Swiss"	1221 "Volga"	1640 "Soviet	2000)"
Islander,	0920 "Suisse	1222 "German	Union, nec"	2060 "Galician
Peloponnesian"	(1980)"	from Russia (1990-	1650 "Tatar (1990-	(1980)"
0490 "Icelander"	0921 "Suisse	2000); German	2000)"	2061 "Gallego
0500 "Irish,	(1990-2000)"	Russian (ACS)"	1651 "Tatar	(1990-2000)"
various subheads,"	0922 "Switzer	1230 "Gruzia	(1980)"	2062 "Galician
0501 "Celtic"	(1990-2000)"	(1990-2000)"	1652 "Crimean	(1990-2000)"
0502 "Irish	0950 "Romansch	1240 "Rom"	(1980)"	2100 "Mexican"
Scotch"	(1980)"	1250 "Hungarian"	1653 "Tuvanian	2101 "Mexican
0510 "Italian	0951 "Romansch	1260 "Magyar"	(1990-2000)"	(1990-2000, ACS,
(1980)"	(1990-2000)"	1280 "Latvian"	1654 "Soviet	PRCS)"
0511 "Italian	0952 "Ladin	1290 "Lithuanian"	Central Asian	2102
(1990-2000, ACS,	(1990-2000)"	1300 "Macedonian"	"Mexicano/Mexican	
PRCS)"	0960 "Suisse	1320 "North	a (1990-2000, ACS,	
0512 "Trieste	Romane (1990-	Caucasian"	PRCS)"	
(1990-2000)"	2000)"	1330 "North	2103 "Mexican	
0513 "San Marino	0961 "Suisse	Caucasian Turkic	Indian"	
(1990-2000)"	Romane (1980)"	(1990-2000)"	2110 "Mexican	
0530 "Abruzzi"	0962 "Ticino"	1400 "Ossetian"	American"	
0540 "Apulian"	0970 "Welsh"	1420 "Polish"	2111 "Mexican	
0550 "Basilicata,	0980 "Scandinavian,	1430 "Kashubian"	American Indian"	
Lucanian"	Nordic"	1440 "Romanian	2130 "Chicano/Chicana"	
0560 "Calabrian"	1000 "Albanian"	(1990-2000, ACS,		
0570 "Amalfin"	1010 "Azerbaijani"	PRCS)"	2180 "Nuevo	
0580 "Emilia	1020 "Belourussian"	1441 "Romanian	Mexicano"	
Romagna"	1030 "Bulgarian"	(1980)"	2181 "Nuevo	
0590 "Rome"	1050 "Carpathian"	1442 "Transylvanian"	Mexicano (1990-	
0600 "Ligurian"	1051 "Carpatho	1450 "Bessarabian	2000)"	
0610 "Lombardian"	Rusyn"	(1980)"	2182 "La Raza	
0620 "Marches"	1052 "Rusyn"	1451 "Bessarabian	(1990-2000)"	
0630 "Molise"	1080 "Cossack	(1990-2000)"	2183 "Mexican	
0640 "Neapolitan"	(1990-2000)"	1452 "Bucovina"	state (1990-2000,	
0650 "Piedmontese"	1081 "Cossack	1460 "Moldavian"	ACS, PRCS)"	
0660 "Puglia"	(1980)"	1470 "Wallachian"	2184 "Tejano/Tejana	
0670 "Sardinian"	1082 "Turkestani	1480 "Russian"	(1990-2000)"	
0680 "Sicilian"	(1990-2000, 2012	1500 "Muscovite"	2190 "Californio"	
0690 "Tuscan"	ACS)"	1520 "Serbian	2210 "Costa	
0700 "Trentino"	1083 "Kirghiz	(1980)"	Rican"	
0710 "Umbrian"	(1980)"	1521 "Serbian	2220 "Guatemalan"	
0720 "Valle	1084 "Turcoman	(1990-2000, ACS,	2230 "Honduran"	
d'Aosta"	(1980)"	PRCS)"	2240 "Nicaraguian"	
0730 "Venetian"	1090 "Croatian"	1522 "Bosnian	2250 "Panamanian	
0750 "Lapp"	1110 "Czechoslovakian"	(1990),	(1980)"	
		Herzegovinian		

2251 "Panamanian (1990-2000, ACS, PRCS)"	3150 "Trinidadian"	3353 "Arawak (1980)"	4700 "South Yemeni"	5570 "Tiv (1980)"
2252 "Canal Zone (1990-2000)"	3160 "Tobagonian"	3360 "Haitian"	4710 "Aden"	5571 "Tiv (1990-2000)"
2260 "Salvadoran"	3170 "U.S. Virgin Islander (1980)"	3370 "Other West Indian"	4800 "United Arab Emirates"	5572 "Yoruba (1990-2000)"
2270 "Latin American (1980)"	3171 "U.S. Virgin Islander (1990-2000)"	3600 "Brazilian"	4820	5610 "Rwandan"
2271 "Central American (1990-2000, ACS, PRCS)"	3172 "St. Croix Islander (1990-2000)"	3650 "San Andres"	"Assyrian/Chaldean /Syriac (1990-2000)"	5640 "Senegalese"
2272 "Latin American (1990-2000, ACS, PRCS)"	3173 "St. John Islander (1990-2000)"	3700 "Guyanese/British Guiana"	4821 "Assyrian "	5660 "Sierra Leonean"
2273 "Latino/Latina (1990-2000, ACS, PRCS)"	3174 "St. Thomas Islander (1990-2000)"	3750 "Providencia"	4822 "Syriac (1980, 2000)"	5680 "Somalian"
2274 "Latin (1990-2000, ACS, PRCS)"	3210 "British Virgin Islander (1980)"	3800 "Surinam/Dutch Guiana"	4823 "Chaldean (2000, ACS, PRCS)"	5690 "Swaziland"
2310 "Argentinean"	3211 "British Virgin Islander	4000 "Algerian"	4900 "Middle Eastern"	5700 "South African"
2320 "Bolivian"	(1990-2000)"	4020 "Egyptian"	4950 "Arab"	5720 "Afrikaner"
2330 "Chilean"	3212 "Antigua (1990-2000)"	4040 "Libyan"	4951 "Arabic	5730 "Natalian"
2340 "Colombian"	(1990-2000, ACS, PRCS)"	4060 "Moroccan (1990-2000, ACS, PRCS)"	(1990-2000, ACS, PRCS)"	5740 "Zulu"
2350 "Ecuadorian"	3220 "British West Indian"	4080 "Tunisian"	4960 "Other Arab"	5760 "Sudanese"
2360 "Paraguayan"	4110 "North African"	4110 "North African"	5000 "Angolan"	5770 "Dinka"
2370 "Peruvian"	3230 "Turks and Caicos Islander"	4120 "Alhucemas"	5020 "Benin"	5780 "Nuer"
2380 "Uruguayan"	4130 "Berber"	4130 "Berber"	5040 "Botswana"	5790 "Fur"
2390 "Venezuelan"	3240 "Anguilla Islander (1980)"	4140 "Rio de Oro"	5060 "Burundian"	5800 "Baggara"
2480 "South American (1980)"	3241 "Anguilla Islander (1990-2000)"	4150 "Bahraini"	5080	5820 "Tanzanian"
2481 "South American (1990-2000, ACS, PRCS)"	3242 "Montserrat Islander (1990-2000)"	4160 "Iranian"	"Cameroonian"	5830 "Tanganyikan"
2482 "Criollo/Criolla (1990-2000)"	3243 "Kirks/Nevis Islander (1990-2000)"	4170 "Iraqi"	5100 "Cape Verdean"	5840 "Zanzibar"
2610 "Puerto Rican"	4190 "Israeli"	4190 "Israeli"	5120 "Central African Republic"	5860 "Togo"
2710 "Cuban"	3244 "St. Christopher (1980)"	4210 "Jordanian"	5130 "Chadian"	5880 "Ugandan"
2750 "Dominican"	3245 "St Vincent Islander (1990);	4220 "Transjordan"	5150 "Congolese"	5890 "Upper Voltan"
2900 "Hispanic"	Vincent-Grenadine	4230 "Kuwaiti"	5160 "Congo-Brazzaville"	5900 "Voltan"
2910 "Spanish"	Islander (2000)	4250 "Lebanese"	5190 "Djibouti"	5910 "Zairian"
2950 "Spanish American"	Census, 2005 ACS, 2005 PRCS"	4270 "Saudi Arabian"	5200 "Equatorial Guinea"	5920 "Zambian"
2960 "Other Spanish/Hispanic"	3280 "Dominica Islander"	4290 "Syrian (1980)"	5210 "Corsico Islander"	5930 "Zimbabwean"
3000 "Bahamian"	4293 "Jebel Druse (1980)"	4310 "Armenian"	5220 "Ethiopian"	5940 "African Islands (1980)"
3010 "Barbadian"	3290 "Grenadian"	4310 "Armenian"	5230 "Eritrean"	5942 "Mauritius (1990-2000)"
3020 "Belizean"	3310 "St Lucia Islander"	4340 "Turkish"	5250 "Gabonese"	5950 "Other Subsaharan Africa"
3030 "Bermudan"	4350 "Yemeni"	4350 "Yemeni"	5270 "Gambian"	5960 "Central African, Middle Congo"
3040 "Cayman Islander"	3320 "French West Indian"	4360 "Omani"	5290 "Ghanian"	5970 "East African"
3080 "Jamaican"	3330 "Guadeloupe Islander"	4370 "Muscat"	5300 "Guinean"	5980 "West African"
3100 "Dutch West Indies"	4380 "Trucial Oman"	4390 "Qatar"	5310 "Guinea Bissau"	5990 "African"
3110 "Aruba Islander"	3340 "Cayenne"	4410 "Bedouin"	5320 "Ivory Coast"	6000 "Afghan"
3120 "St Maarten Islander"	3350 "West Indian (1990-2000, ACS, PRCS)"	4420 "Kurdish"	5340 "Kenyan"	6010 "Baluchi"
3140 "Trinidadian/Tobagonian"	3351 "West Indian (1980)"	4440 "Kuria Muria Islander"	5380 "Lesotho"	6020 "Pathan"
	3352 "Caribbean (1980)"	4650 "Palestinian"	5410 "Liberian"	6030 "Bengali (1980)"
		4660 "Gazan"	5430	6031 "Bangladeshi (1990-2000, ACS, PRCS)"
		4670 "West Bank"	5450 "Malawian"	6070 "Bhutanese"
			5460 "Malian"	6090 "Nepali"
			5470 "Mauritanian"	
			5490 "Mozambican"	
			5500 "Namibian"	
			5510 "Niger"	
			5530 "Nigerian"	
			5540 "Fulani"	
			5550 "Hausa"	
			5560 "Ibo"	

6150 "Asian Indian (1980)"	7071 "Cantonese (1990-2000, ACS, PRCS)"	7740 "Singaporean" 7760 "Thai" 7770 "Black Thai" 7780 "Western Lao" 7820 "Taiwanese" 7850 "Vietnamese, Annamese" 7860 "Katu" 7870 "Ma" 7880 "Mnong" 7900 "Montagnard" 7920 "Indochinese" 7930 "Eurasian" 7931 "Amerasian (1990-2000, ACS, PRCS)"	8180 "Tahitian, French Polynesian, Society Islander" 8190 "Niuean" 8200 "Micronesian (1990-2000, ACS, PRCS)" 8201 "Micronesian (1980)" 8202 "U.S. Trust Terr of the Pacific (1980)" 8210 "Guamanian" 8220 "Chamorro Islander" 8230 "Saipanese (1990-2000)" 8231 "Saipanese (1980)" 8232 "Norther Marianas (1980)" 8240 "Palauan" 8250 "Marshall Islander" 8260 "Kosraean" 8270 "Ponapean (1990-2000)" 8271 "Ponapean (1980)" 8272 "Mokilese (1980)" 8273 "Ngatikese (1980)" 8274 "Pingelapese (1980)" 8280 "Chuukese (1990-2000)" 8281 "Hall Islander (1980)" 8282 "Mortlockese (1980)" 8283 "Kapinagamaranga n (1990-2000)" 8091 "Kapinagamaranga n (1980)" 8092 "Nukuoroan (1980)" 8093 "Nisei (1990- 2000)" 7402 "Issei (1990- 2000)" 7403 "Nisei (1990- 2000)" 7404 "Sansei (1990-2000)" 7405 "Yonsei (1990-2000)" 7406 "Gosei (1990-2000)" 7407 "Singhalese" (1990-2000)" 7408 "Veddah" 7409 "Maldivian" 7410 "Burmese (1990-2000, ACS, PRCS)" 7411 "Kashmiri (1990-2000)" 7412 "Okinawan" 7500 "Burmese (1980)" 7501 "Korean" 7650 "Laotian" 7660 "Meo" 7680 "Hmong" 7700 "Malaysian (1980)" 7701 "Malaysian (1990-2000, ACS, PRCS)" 7702 "North Borneo (1990- 2000)"	8330 "Tarawa Islander (1990- 2000)" 8340 "Tinian Islander (1990- 2000)" 8400 "Melanesian Islander" 8410 "Fijian" 8430 "New Guinean" 8440 "Papuan" 8450 "Solomon Islander" 8460 "New Caledonian Islander" 8470 "Vanuatuun" 8500 "Pacific Islander (1990- 2000, ACS, PRCS)" 8501 "Campbell Islander (1980)" 8502 "Christmas Islander (1980)" 8503 "Kermadec Islander (1980)" 8504 "Midway Islander (1980)" 8505 "Phoenix Islander (1980)" 8506 "Wake Islander (1980)" 8600 "Oceania" 8620 "Chamolinian (1990-2000)" 8630 "Reserved Codes" 8700 "Other Pacific" 9000 "Afro- American" 9001 "Afro- American (1990- 2000, ACS, PRCS)" 9002 "Black (1990-2000, ACS, PRCS)" 9003 "Negro (1990-2000, ACS, PRCS)" 9004 "Nonwhite (1990-2000)" 9005 "Colored (1990-2000)" 9006 "Creole (1990-2000, ACS, PRCS)" 9007 "Mulatto (1990-2000)" 9008 "Afro" 9020 "African- American (1990- 2000, ACS, PRCS)"
-------------------------------	---	--	--	--

9130 "Central American Indian (1990-2000, ACS, PRCS)"	9420 "Alaska"	9960 "Uncodable"
9140 "South American Indian (1990-2000, ACS, PRCS)"	9430 "Arizona"	9961 "Not Classified"
9200 "American Indian (all tribes)"	9440 "Arkansas"	9980 "Other"
9201 "American Indian-English-French"	9450 "California"	9990 "Not Reported"
9202 "American Indian-English-German"	9460 "Colorado"	
9203 "American Indian-English-Irish"	9470 "Connecticut"	
9204 "American Indian-German-Irish"	9480 "District of Columbia"	
9205 "Cherokee"	9490 "Delaware"	
9206 "Native American"	9500 "Florida"	
9207 "Indian"	9510 "Georgia"	
9210 "Aleut"	9520 "Idaho"	
9220 "Eskimo"	9530 "Illinois"	
9230 "Inuit"	9540 "Indiana"	
9240 "White/Caucasian"	9550 "Iowa"	
9241 "White/Caucasian (1990-2000, ACS, PRCS)"	9560 "Kansas"	
9242 "Anglo (1990-2000, ACS, PRCS)"	9570 "Kentucky"	
9243 "Appalachian (1990-2000, ACS, PRCS)"	9580 "Louisiana"	
9244 "Aryan (1990-2000)"	9590 "Maine"	
9300 "Greenlander"	9600 "Maryland"	
9310 "Canadian"	9610 "Massachusetts"	
9330 "Newfoundland"	9620 "Michigan"	
9340 "Nova Scotian"	9630 "Minnesota"	
9350 "French Canadian"	9640 "Mississippi"	
9360 "Acadian"	9650 "Missouri"	
9361 "Acadian (1990-2000, ACS, PRCS)"	9660 "Montana"	
9362 "Cajun (1990-2000, ACS, PRCS)"	9670 "Nebraska"	
9390 "American"	9680 "Nevada"	
9391 "American/United States"	9690 "New Hampshire"	
9400 "United States"	9700 "New Jersey"	
9410 "Alabama"	9710 "New Mexico"	
	9720 "New York"	
	9730 "North Carolina"	
	9740 "North Dakota"	
	9750 "Ohio"	
	9760 "Oklahoma"	
	9770 "Oregon"	
	9780 "Pennsylvania"	
	9790 "Rhode Island"	
	9800 "South Carolina"	
	9810 "South Dakota"	
	9820 "Tennessee"	
	9830 "Texas"	
	9840 "Utah"	
	9850 "Vermont"	
	9860 "Virginia"	
	9870 "Washington"	
	9880 "West Virginia"	
	9890 "Wisconsin"	
	9900 "Wyoming"	
	9930 "Southerner"	
	9940 "North American"	
	9950 "Mixture"	