

Real data analysis

We conducted a real data analysis using the Household Pulse Survey (HPS) Phase 4.2 Cycle 09 data collected by United States Census Bureau in 2024

(<https://www.census.gov/programs-surveys/household-pulse-survey/data/tables.html>).

HPS provides a rich resource to study how emergent issues are impacting U.S. households from a social and economic perspective. There are 220 variables and 51,280 records in the downloaded data table. We picked 3 variables as outcome and 16 variables as predictors in the real data analysis. After removing records missing any variable, there are 42,492 records in the complete dataset. Most of the variables are categorical, while there are a few continuous variables. The three outcome variables are Anxious status (43.8% Yes), Trouble paying energy bill (26.1% Yes), and Perception of price changes (90.4% Yes). The predictor variables contain age, gender, race, education, marital status, income, number of adults/kids in the household, employment status, limitations in seeing, hearing, mobility, remembering, selfcare, understanding, and social support. The overall and stratified distribution by outcome variables of predictors were summarized in Table 1-3. Wilcoxon and Chi-square test were used to test the univariate association between predictor and outcome variable. The missing rate in HPS data table is low (<10%) for most of the variables. We manually generated the response index using

$$\text{Logit}(p) = -1.14 + \text{Race}^{(1/3)} * \log_{10}(\text{Gender} + \text{Number of kids}) + \sin(\text{Gender} * \text{Number of kids} + 0.2 * \text{Income}) + \text{Race}^{(1/2)} / \text{Income}$$

which resulted in ~50% missing rate.

We applied the previously discussed methods and procedures in estimating the proportion of Yes in the outcome variables, including LM, GAM, KNN, RF, RT, SVM, XGB, and DL. A random sample of 1000 records were drawn from the complete dataset and used in the analysis. The estimates from various methods were evaluated against the value in the complete dataset to calculate the relative bias. The result is shown in Table 4. The computational code is available at: <https://github.com/xu1912/ImputationMethods>.

Table 1. Predictors by Anxious (Y/N) status

	Total (N=42492)	N (N=23882)	Y (N=18610)	P-value
Age				
Mean (SD)	55 (\pm 16)	59 (\pm 15)	49 (\pm 15)	<0.001*
Median [Q1, Q3]	56 [42, 68]	62 [48, 71]	48 [37, 62]	
Gender				
Male	19391 (45.6%)	12086 (50.6%)	7305 (39.3%)	<0.001
Female	23101 (54.4%)	11796 (49.4%)	11305 (60.7%)	
Race				
White	35335 (83.2%)	19691 (82.5%)	15644 (84.1%)	<0.001
Black	3120 (7.3%)	1952 (8.2%)	1168 (6.3%)	
Asian	1824 (4.3%)	1185 (5.0%)	639 (3.4%)	
Other	2213 (5.2%)	1054 (4.4%)	1159 (6.2%)	
Education				
Graduate degree	10856 (25.5%)	6369 (26.7%)	4487 (24.1%)	<0.001
Bachelor's degree	11945 (28.1%)	6556 (27.5%)	5389 (29.0%)	
Associate's degree	4510 (10.6%)	2550 (10.7%)	1960 (10.5%)	
Some college	8810 (20.7%)	4733 (19.8%)	4077 (21.9%)	
High school graduate	5516 (13.0%)	3212 (13.4%)	2304 (12.4%)	

	Total (N=42492)	N (N=23882)	Y (N=18610)	P-value
Some high school	591 (1.4%)	316 (1.3%)	275 (1.5%)	
Less than high-school	264 (0.6%)	146 (0.6%)	118 (0.6%)	
Marital status				
Now married	23779 (56.0%)	14472 (60.6%)	9307 (50.0%)	<0.001
Widowed	2851 (6.7%)	1906 (8.0%)	945 (5.1%)	
Divorced	7221 (17.0%)	3957 (16.6%)	3264 (17.5%)	
Separated	692 (1.6%)	320 (1.3%)	372 (2.0%)	
Never married	7949 (18.7%)	3227 (13.5%)	4722 (25.4%)	
Number of adults in the household				
Mean (SD)	2.0 (± 0.91)	2.0 (± 0.88)	2.0 (± 0.95)	0.012*
Median [Q1, Q3]	2.0 [1.0, 2.0]	2.0 [1.0, 2.0]	2.0 [1.0, 2.0]	
Number of kids in the household				
Mean (SD)	0.51 (± 0.97)	0.44 (± 0.91)	0.60 (± 1.0)	<0.001*
Median [Q1, Q3]	0 [0, 1.0]	0 [0, 0]	0 [0, 1.0]	
Employment status				
Yes	24947 (58.7%)	12906 (54.0%)	12041 (64.7%)	<0.001
No	17545 (41.3%)	10976 (46.0%)	6569 (35.3%)	
Limitations in seeing				

	Total (N=42492)	N (N=23882)	Y (N=18610)	P-value
N	29819 (70.2%)	18518 (77.5%)	11301 (60.7%)	<0.001
Y	12673 (29.8%)	5364 (22.5%)	7309 (39.3%)	
Limitations in hearing				
N	34298 (80.7%)	19676 (82.4%)	14622 (78.6%)	<0.001
Y	8194 (19.3%)	4206 (17.6%)	3988 (21.4%)	
Limitations in mobility				
N	31413 (73.9%)	18395 (77.0%)	13018 (70.0%)	<0.001
Y	11079 (26.1%)	5487 (23.0%)	5592 (30.0%)	
Limitations in remembering				
N	29615 (69.7%)	19652 (82.3%)	9963 (53.5%)	<0.001
Y	12877 (30.3%)	4230 (17.7%)	8647 (46.5%)	
Limitations in self-care				
N	39525 (93.0%)	23126 (96.8%)	16399 (88.1%)	<0.001
Y	2967 (7.0%)	756 (3.2%)	2211 (11.9%)	
Limitations in understanding or being understood				
N	40153 (94.5%)	23301 (97.6%)	16852 (90.6%)	<0.001
Y	2339 (5.5%)	581 (2.4%)	1758 (9.4%)	

	Total (N=42492)	N (N=23882)	Y (N=18610)	P-value
Social and emotional support				
Always	12574 (29.6%)	10131 (42.4%)	2443 (13.1%)	<0.001
Usually	15633 (36.8%)	8843 (37.0%)	6790 (36.5%)	
Sometimes	7811 (18.4%)	2477 (10.4%)	5334 (28.7%)	
Rarely	3816 (9.0%)	917 (3.8%)	2899 (15.6%)	
Never	2658 (6.3%)	1514 (6.3%)	1144 (6.1%)	
Income				
<\$25,000	4750 (11.2%)	2125 (8.9%)	2625 (14.1%)	<0.001
\$25,000 - \$34,999	3502 (8.2%)	1878 (7.9%)	1624 (8.7%)	
\$35,000 - \$49,999	4468 (10.5%)	2446 (10.2%)	2022 (10.9%)	
\$50,000 - \$74,999	6942 (16.3%)	3936 (16.5%)	3006 (16.2%)	
\$75,000 - \$99,999	5804 (13.7%)	3384 (14.2%)	2420 (13.0%)	
\$100,000 - \$149,999	7567 (17.8%)	4370 (18.3%)	3197 (17.2%)	
\$150,000 - \$199,999	4099 (9.6%)	2456 (10.3%)	1643 (8.8%)	
\$200,000 and above	5360 (12.6%)	3287 (13.8%)	2073 (11.1%)	

*Wilcoxon test used when continuous variable failed in Shapiro normality test at level of 0.05.

Table 2. Predictors by Trouble paying energy bill (Y/N)

	Total (N=42492)	N (N=31392)	Y (N=11100)	P-value
Age				
Mean (SD)	55 (\pm 16)	56 (\pm 16)	52 (\pm 14)	<0.001*
Median [Q1, Q3]	56 [42, 68]	58 [43, 69]	52 [41, 63]	
Gender				
Male	19391 (45.6%)	15114 (48.1%)	4277 (38.5%)	<0.001
Female	23101 (54.4%)	16278 (51.9%)	6823 (61.5%)	
Race				
White	35335 (83.2%)	26714 (85.1%)	8621 (77.7%)	<0.001
Black	3120 (7.3%)	1889 (6.0%)	1231 (11.1%)	
Asian	1824 (4.3%)	1474 (4.7%)	350 (3.2%)	
Other	2213 (5.2%)	1315 (4.2%)	898 (8.1%)	
Education				
Graduate degree	10856 (25.5%)	9411 (30.0%)	1445 (13.0%)	<0.001
Bachelor's degree	11945 (28.1%)	9641 (30.7%)	2304 (20.8%)	
Associate's degree	4510 (10.6%)	2961 (9.4%)	1549 (14.0%)	
Some college	8810 (20.7%)	5647 (18.0%)	3163 (28.5%)	
High school graduate	5516 (13.0%)	3328 (10.6%)	2188 (19.7%)	

	Total (N=42492)	N (N=31392)	Y (N=11100)	P-value
Some high school	591 (1.4%)	276 (0.9%)	315 (2.8%)	
Less than high-school	264 (0.6%)	128 (0.4%)	136 (1.2%)	
Marital status				
Now married	23779 (56.0%)	18752 (59.7%)	5027 (45.3%)	<0.001
Widowed	2851 (6.7%)	2112 (6.7%)	739 (6.7%)	
Divorced	7221 (17.0%)	4727 (15.1%)	2494 (22.5%)	
Separated	692 (1.6%)	363 (1.2%)	329 (3.0%)	
Never married	7949 (18.7%)	5438 (17.3%)	2511 (22.6%)	
Number of adults in the household				
Mean (SD)	2.0 (± 0.91)	2.0 (± 0.84)	2.1 (± 1.1)	<0.001*
Median [Q1, Q3]	2.0 [1.0, 2.0]	2.0 [1.0, 2.0]	2.0 [1.0, 3.0]	
Number of kids in the household				
Mean (SD)	0.51 (± 0.97)	0.46 (± 0.91)	0.66 (± 1.1)	<0.001*
Median [Q1, Q3]	0 [0, 1.0]	0 [0, 1.0]	0 [0, 1.0]	
Employment status				
Yes	24947 (58.7%)	18381 (58.6%)	6566 (59.2%)	0.275
No	17545 (41.3%)	13011 (41.4%)	4534 (40.8%)	

	Total (N=42492)	N (N=31392)	Y (N=11100)	P-value
Limitations in seeing				
N	29819 (70.2%)	23780 (75.8%)	6039 (54.4%)	<0.001
Y	12673 (29.8%)	7612 (24.2%)	5061 (45.6%)	
Limitations in hearing				
N	34298 (80.7%)	25772 (82.1%)	8526 (76.8%)	<0.001
Y	8194 (19.3%)	5620 (17.9%)	2574 (23.2%)	
Limitations in mobility				
N	31413 (73.9%)	24549 (78.2%)	6864 (61.8%)	<0.001
Y	11079 (26.1%)	6843 (21.8%)	4236 (38.2%)	
Limitations in remembering				
N	29615 (69.7%)	23378 (74.5%)	6237 (56.2%)	<0.001
Y	12877 (30.3%)	8014 (25.5%)	4863 (43.8%)	
Limitations in self-care				
N	39525 (93.0%)	29919 (95.3%)	9606 (86.5%)	<0.001
Y	2967 (7.0%)	1473 (4.7%)	1494 (13.5%)	
Limitations in understanding or being understood				

	Total (N=42492)	N (N=31392)	Y (N=11100)	P-value
N	40153 (94.5%)	30149 (96.0%)	10004 (90.1%)	<0.001
Y	2339 (5.5%)	1243 (4.0%)	1096 (9.9%)	
Social and emotional support				
Always	12574 (29.6%)	10426 (33.2%)	2148 (19.4%)	<0.001
Usually	15633 (36.8%)	12524 (39.9%)	3109 (28.0%)	
Sometimes	7811 (18.4%)	5021 (16.0%)	2790 (25.1%)	
Rarely	3816 (9.0%)	1926 (6.1%)	1890 (17.0%)	
Never	2658 (6.3%)	1495 (4.8%)	1163 (10.5%)	
Income				
<\$25,000	4750 (11.2%)	2317 (7.4%)	2433 (21.9%)	<0.001
\$25,000 - \$34,999	3502 (8.2%)	1952 (6.2%)	1550 (14.0%)	
\$35,000 - \$49,999	4468 (10.5%)	2804 (8.9%)	1664 (15.0%)	
\$50,000 - \$74,999	6942 (16.3%)	4818 (15.3%)	2124 (19.1%)	
\$75,000 - \$99,999	5804 (13.7%)	4439 (14.1%)	1365 (12.3%)	
\$100,000 - \$149,999	7567 (17.8%)	6270 (20.0%)	1297 (11.7%)	
\$150,000 - \$199,999	4099 (9.6%)	3681 (11.7%)	418 (3.8%)	
\$200,000 and above	5360 (12.6%)	5111 (16.3%)	249 (2.2%)	

*Wilcoxon test used when continuous variable failed in Shapiro normality test at level of 0.05.

Table 3. Predictors by Perception of price changes (Y/N)

	Total (N=42492)	N (N=4058)	Y (N=38434)	P-value
Age				
Mean (SD)	55 (±16)	59 (±15)	55 (±16)	<0.001*
Median [Q1, Q3]	56 [42, 68]	62 [48, 71]	55 [42, 68]	
Gender				
Male	19391 (45.6%)	2413 (59.5%)	16978 (44.2%)	<0.001
Female	23101 (54.4%)	1645 (40.5%)	21456 (55.8%)	
Race				
White	35335 (83.2%)	3511 (86.5%)	31824 (82.8%)	<0.001
Black	3120 (7.3%)	267 (6.6%)	2853 (7.4%)	
Asian	1824 (4.3%)	142 (3.5%)	1682 (4.4%)	
Other	2213 (5.2%)	138 (3.4%)	2075 (5.4%)	
Education				
Graduate degree	10856 (25.5%)	1667 (41.1%)	9189 (23.9%)	<0.001
Bachelor's degree	11945 (28.1%)	1264 (31.1%)	10681 (27.8%)	
Associate's degree	4510 (10.6%)	271 (6.7%)	4239 (11.0%)	
Some college	8810 (20.7%)	533 (13.1%)	8277 (21.5%)	
High school graduate	5516 (13.0%)	272 (6.7%)	5244 (13.6%)	

	Total (N=42492)	N (N=4058)	Y (N=38434)	P-value
Some high school	591 (1.4%)	29 (0.7%)	562 (1.5%)	
Less than high-school	264 (0.6%)	22 (0.5%)	242 (0.6%)	
Marital status				
Now married	23779 (56.0%)	2564 (63.2%)	21215 (55.2%)	<0.001
Widowed	2851 (6.7%)	267 (6.6%)	2584 (6.7%)	
Divorced	7221 (17.0%)	577 (14.2%)	6644 (17.3%)	
Separated	692 (1.6%)	31 (0.8%)	661 (1.7%)	
Never married	7949 (18.7%)	619 (15.3%)	7330 (19.1%)	
Number of adults in the household				
Mean (SD)	2.0 (± 0.91)	1.9 (± 0.80)	2.0 (± 0.92)	<0.001*
Median [Q1, Q3]	2.0 [1.0, 2.0]	2.0 [1.0, 2.0]	2.0 [1.0, 2.0]	
Number of kids in the household				
Mean (SD)	0.51 (± 0.97)	0.37 (± 0.82)	0.53 (± 0.98)	<0.001*
Median [Q1, Q3]	0 [0, 1.0]	0 [0, 0]	0 [0, 1.0]	
Employment status				
Yes	24947 (58.7%)	2231 (55.0%)	22716 (59.1%)	<0.001
No	17545 (41.3%)	1827 (45.0%)	15718 (40.9%)	

	Total (N=42492)	N (N=4058)	Y (N=38434)	P-value
Limitations in seeing				
N	29819 (70.2%)	3351 (82.6%)	26468 (68.9%)	<0.001
Y	12673 (29.8%)	707 (17.4%)	11966 (31.1%)	
Limitations in hearing				
N	34298 (80.7%)	3432 (84.6%)	30866 (80.3%)	<0.001
Y	8194 (19.3%)	626 (15.4%)	7568 (19.7%)	
Limitations in mobility				
N	31413 (73.9%)	3381 (83.3%)	28032 (72.9%)	<0.001
Y	11079 (26.1%)	677 (16.7%)	10402 (27.1%)	
Limitations in remembering				
N	29615 (69.7%)	3325 (81.9%)	26290 (68.4%)	<0.001
Y	12877 (30.3%)	733 (18.1%)	12144 (31.6%)	
Limitations in self-care				
N	39525 (93.0%)	3928 (96.8%)	35597 (92.6%)	<0.001
Y	2967 (7.0%)	130 (3.2%)	2837 (7.4%)	
Limitations in understanding or being understood				
N	40153 (94.5%)	3950 (97.3%)	36203 (94.2%)	<0.001

	Total (N=42492)	N (N=4058)	Y (N=38434)	P-value
Y	2339 (5.5%)	108 (2.7%)	2231 (5.8%)	
Social and emotional support				
Always	12574 (29.6%)	1698 (41.8%)	10876 (28.3%)	<0.001
Usually	15633 (36.8%)	1561 (38.5%)	14072 (36.6%)	
Sometimes	7811 (18.4%)	449 (11.1%)	7362 (19.2%)	
Rarely	3816 (9.0%)	169 (4.2%)	3647 (9.5%)	
Never	2658 (6.3%)	181 (4.5%)	2477 (6.4%)	
Income				
<\$25,000	4750 (11.2%)	238 (5.9%)	4512 (11.7%)	<0.001
\$25,000 - \$34,999	3502 (8.2%)	194 (4.8%)	3308 (8.6%)	
\$35,000 - \$49,999	4468 (10.5%)	268 (6.6%)	4200 (10.9%)	
\$50,000 - \$74,999	6942 (16.3%)	488 (12.0%)	6454 (16.8%)	
\$75,000 - \$99,999	5804 (13.7%)	497 (12.2%)	5307 (13.8%)	
\$100,000 - \$149,999	7567 (17.8%)	767 (18.9%)	6800 (17.7%)	
\$150,000 - \$199,999	4099 (9.6%)	542 (13.4%)	3557 (9.3%)	
\$200,000 and above	5360 (12.6%)	1064 (26.2%)	4296 (11.2%)	

*Wilcoxon test used when continuous variable failed in Shapiro normality test at level of 0.05.

Table 4. Relative bias (RB) (%) of different methods for the three outcome variables

Model	Method	Anxious	Trouble paying energy bill	Perception of price changes
LM	IM	1.38	0.67	1.22
	PS	2.28	2.37	1.45
	DML	1.92	0.79	0.91
GAM	IM	1.38	0.67	1.22
	PS	2.28	2.37	1.45
	DML	2.04	1.15	1.26
KNN	IM	2.21	5.09	2.14
	PS	13.81	13.21	17.15
	DML	19.30	41.71	3.30
XGB	IM	2.95	0.54	21.45
	PS	7.77	11.78	5.69
	DML	0.95	2.84	1.01
RT	IM	2.21	0.46	1.22
	PS	2.55	4.77	0.73
	DML	1.25	2.41	0.73
SVM	IM	0.52	1.81	1.22
	PS	1.29	2.22	3.81
	DML	0.53	1.81	1.11
RF	IM	0.76	2.50	3.92
	PS	23.91	22.91	25.74
	DML	2.79	2.22	1.93
DL (0.01)	IM	5.65	8.26	2.84
	PS	5.78	10.64	6.86
	DML	1.81	4.41	0.85
DL (0.001)	IM	6.79	5.71	2.83
	PS	8.20	13.75	8.49
	DML	1.02	2.49	1.03