Homework 1

Research Methods, Spring 2024

Answer Key

My answers to the homework questions are described below. Note that I do the analysis for these answers in a separate `R` script. You can read in the full data as part of your markdown file, but that takes some time to compile to pdf. So I run the analysis separately, save the workspace with only the summary stats, figures, and tables that I need, and then load the workspace in the final qmd. My analysis file is available in the analysis folder. Enjoy!

Enrollment Data

Answer the following based on the enrollment data:

1. How many observations exist in your current dataset?

First we need to create the enrollment data. Working with the Medicare Advantage Github Repository, you should have created a "full.ma.data" object. Then we just count the total number of plans. This yields 19,126,783 total observations in the full dataset, which means there are 19,126,783 unique combinations of contract/plan/county/year.

2. How many different *plan_types* exist in the data?

To do this, we need to group by plan type and count the number of unique plan types. I did this by creating a table of unique plan types (since we'll need this for the next question anyway). The resulting table yields 27 rows, so there are 27 total plan types.

3. Provide a table of the count of plans under each plan type in each year.

See Table 1.

Table 1: Plan types by year

Plan Type	2007	2008	2009	2010	2011	2012	2013	2014	2015
Medicare Prescription Drug Plan	920,058	963,478	945,794	893,609	771,694	815,223	826,907	1,122,209	991,457
PFFS	364,285	630,756	683,361	385,733	45,781	36,423	31,919	24,905	13,658
HMO/HMOPOS	60,012	70,176	479,978	506,802	528,473	507,272	530,909	523,304	479,275
Employer/Union Only Direct Contract PDP	32,358	29,113	25,860	28,700	28,697	28,669	25,526	25,528	25,630
Regional PPO	26,402	27,990	25,943	24,442	22,773	21,602	19,970	19,773	17,578
Local PPO	17,427	38,470	405,197	417,551	515,700	636,701	633,884	664,716	704,993
1876 Cost	5,855	5,459	5,825	6,035	6,851	7,633	7,731	7,069	7,157
MSA	4,422	16,515	12,267	135	6,421	6,416	6,431	6,449	6,518
MSA Demo	3,274	0	0	0	0	0	0	0	0
Employer Direct PFFS	3,247	0	0	0	0	0	0	0	0
SHMO	1,125	0	0	0	0	0	0	0	0
MN Senior Health Options	968	0	0	0	0	0	0	0	0
PSO (State License)	421	535	87	123	176	171	0	0	0
National PACE	405	548	616	717	781	858	953	1,118	1,216
PSO (Federal Waiver of State License)	162	0	0	0	0	0	0	0	0
Continuing Care Retirement Community	95	122	158	142	0	0	0	0	0
ESRD I	75	122	123	117	0	0	0	0	0
MA Health Senior Care Options	73	0	0	0	0	0	0	0	0
WI Partnership Program	42	0	0	0	0	0	0	0	0
MN Disability Health Options	21	0	0	0	0	0	0	0	0
Pilot	15	12	201	53	3	3	2	2	2
HCPP - 1833 Cost	13	13	3,938	3,604	11	11	10	9	9
ESRD II	12	12	7	8	0	0	0	0	0
Employer/Union Only Direct Contract PFFS	0	3,332	3,335	3,332	3,329	3,323	0	0	0
RFB PFFS	0	0	3,006	0	0	0	0	0	0
Medicare-Medicaid Plan HMO/HMOPOS	0	0	0	0	0	0	265	1,319	4,130

4. Remove all special needs plans (SNP), employer group plans (eghp), and all "800-series" plans. Provide an updated table after making these exclusions.

I remove the relevant plans just by applying the relevant filter to the full ma data and then creating the table of plan types. Counts of different plan types with these exclusions are presented in Table 2

Table 2: Revised plan types by year

Plan Type	2007	2008	2009	2010	2011	2012	2013	2014	2015
Medicare Prescription Drug Plan	398,167	428,936	415,027	391,205	295,458	289,044	278,091	301,082	269,153
PFFS	51,987	105,859	89,586	54,119	22,038	17,449	12,945	6,053	4,232
HMO/HMOPOS	30,670	34,545	36,166	34,460	33,931	37,551	37,179	38,893	36,588
Regional PPO	7,254	7,794	8,470	10,659	10,995	11,279	9,660	10,420	8,531
Local PPO	6,116	7,612	9,929	11,652	13,874	17,030	17,089	17,169	16,728
1876 Cost	5,048	4,577	4,781	4,923	5,829	6,647	6,759	6,207	6,329
MSA	2,177	3,303	2,459	68	131	132	145	163	232
SHMO	458	0	0	0	0	0	0	0	0
National PACE	395	548	615	717	781	858	953	1,118	1,216
PSO (State License)	376	394	75	97	141	143	0	0	0
MSA Demo	129	0	0	0	0	0	0	0	0
PSO (Federal Waiver of State License)	110	0	0	0	0	0	0	0	0
ESRD I	75	122	123	117	0	0	0	0	0
Continuing Care Retirement Community	68	66	60	64	0	0	0	0	0
0	0	0	13,619	29,733	0	0	0	0	0
RFB PFFS	0	0	3,006	0	0	0	0	0	0
Medicare-Medicaid Plan HMO/HMOPOS	0	0	0	0	0	0	265	1,319	4,130

5. Merge the the contract service area data to the enrollment data and restrict the data only to contracts that are approved in their respective counties. Limit your dataset only to plans with non-missing enrollment data. Provide a graph showing the average number of Medicare Advantage enrollees per county from 2008 to 2015.

Now we can join that dataset to our MA data. I use an inner join, which means I'm only taking rows that match in both datasets. I then apply the filter to remove plans with missing enrollment data, from which we can form the graph of average enrollments per county, as reflected in Figure 1.

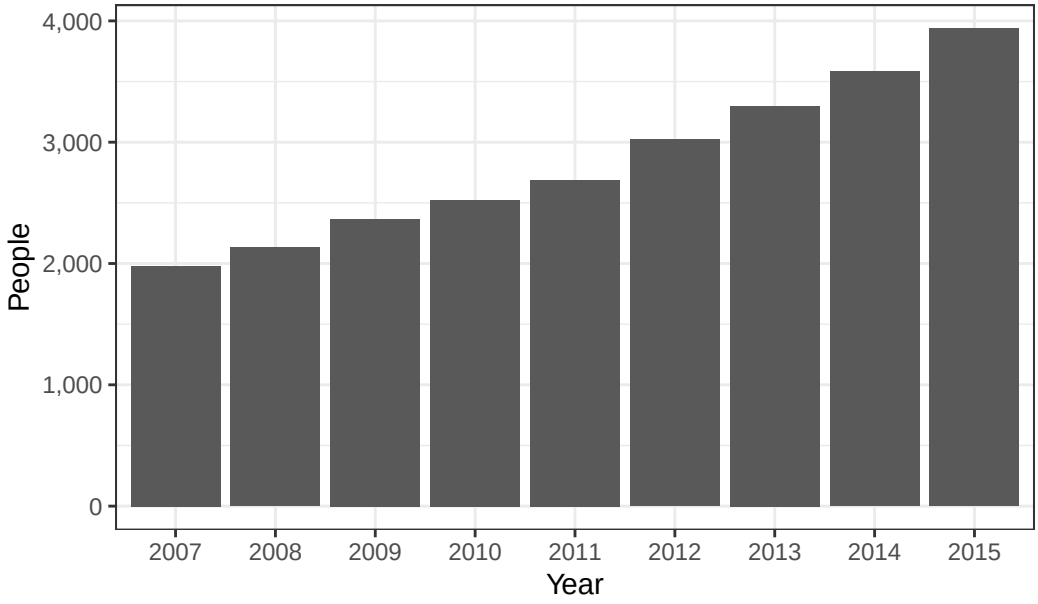

Figure 1: Average Enrollment

Premium Data

6. Merge the plan characteristics data to the dataset you created in Step 5 above. Provide a graph showing the average premium over time. Don't forget about formatting!

As mentioned in the instructions, we first need to merge in the market penetration data to provide a crosswalk between the plan/contract info and the plan characteristics. Next we need to fill in the state information. I do this by creating a table of unique state names and then merging this back to the original data. Finally, we can read in the premium data and merge that information to the final dataset

A graph of average premiums over time is presented in Figure 2. Note the spike in premiums in 2014. What's that?

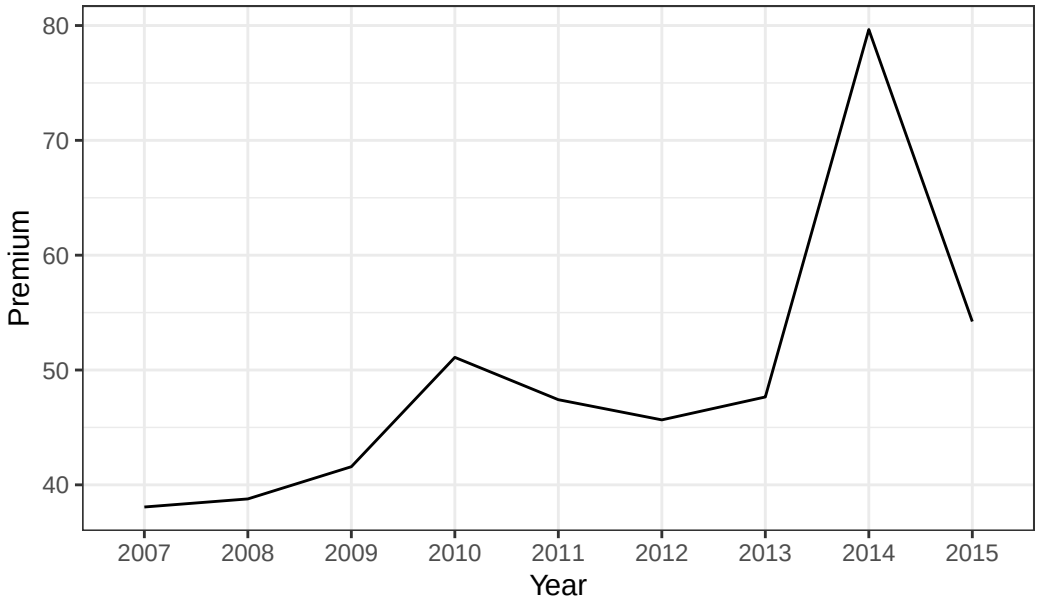

Figure 2: Average Premiums

7. Provide a graph showing the percentage of $0 premium plans over time. Also…remember to format things.

A graph of the percentage of $0 premium plans is in Figure 3. Consistent with Figure 2, we see a large drop (down to 0%) in the percentage of 0 premium plans in 2014. If we also look at the number of missing premium plans, we would see a big spike in 2014. Effectively, these premiums are 0 in some years but listed as missing in 2014.

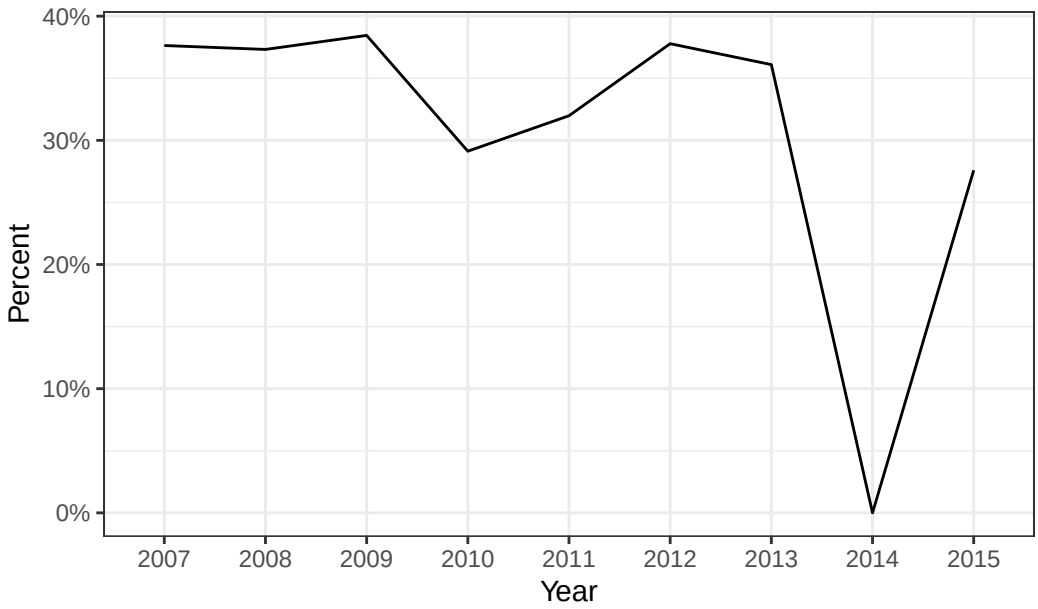

Figure 3: Share of 0 premium plans

Summary Questions

With all of this data work and these great summaries, let's take a step back and think about what all this means.

8. Why did we drop the "800-series" plans?

These are plans that aren't available to all people. There are sometimes referred to as "Employer Group Waiver Plans". Since not everyone has access to these plans, summaries including these plans aren't reflective of an average enrollee's experience in the Medicare Advantage program.

9. Why do so many plans charge a $0 premium? What does that really mean to a beneficiary?

All beneficiaries still pay a Part B premium (nearly $180 in 2022). So a plan with no premium really just means it's a plan with no additional premium in excess of the Part B premium.

10. Briefly describe your experience working with these data (just a few sentences). Tell me one thing you learned and one thing that really aggravated you.

One thing I learned as an instructor is that it takes a couple of days to get all of the kinks out in the workflow process (git, github, r, rstudio, rmarkdown). I think next year I'll have a 1-2 hour tutorial on a weekend or something to make sure everyone has this process in place before we start on the homework. On the bright side, this is the biggest dataset we'll work with all year!