Homework 1 - Research in Health Economics

Alexia Witthaus Viñé

2023-01-25

```
# laod("howk1_workspace.Rdata")
```

```
r  format(tot.obs, big.marks = ",")
```

```
full_ma_data <- read_rds("data/output/full_ma_data.rds")
```

Question 1

```
paste0("There are ", nrow(full_ma_data), " observations in my current dataset")
```

```
## [1] "There are 19126783 observations in my current dataset"
```

Question 2

```
# do -1 because there is one of the 27 is NA, and that doesn't correspond to a plan, ask to do with the
paste0("There are ", length(unique(full_ma_data$plan_type)) -1, " unique plan types in the dataset")
```

```
## [1] "There are 26 unique plan types in the dataset"
```

Question 3

```
# q_3_df <- full_ma_data %>%
#   group_by(year, plan_type)%>%
#   summarize(n_under_plan_type = n())%>%
#   spread(year, n_under_plan_type)

read_rds("results/q_3_df.rds")
```

plan_type	2007	2008	2009	2010	2011	2012	2013	2014	2015
1876 Cost	5855	5459	5825	6035	6851	7633	7731	7069	7157

plan_type	2007	2008	2009	2010	2011	2012	2013	2014	2015
Continuing Care Retirement Community	95	122	158	142	NA	NA	NA	NA	NA
Employer Direct PFFS	3247	NA	NA	NA	NA	NA	NA	NA	NA
Employer/Union Only Direct Contract PDP	32358	29113	25860	28700	28697	28669	25526	25528	25630
Employer/Union Only Direct Contract PFFS	NA	3332	3335	3332	3329	3323	NA	NA	NA
ESRD I	75	122	123	117	NA	NA	NA	NA	NA
ESRD II	12	12	7	8	NA	NA	NA	NA	NA
HCPP - 1833 Cost	13	13	3938	3604	11	11	10	9	9
HMO/HMOPOS	60012	70176	479978	506802	528473	507272	530909	523304	479275
Local PPO	17427	38470	405197	417551	515700	636701	633884	664716	704993
MA Health Senior Care Options	73	NA	NA	NA	NA	NA	NA	NA	NA
Medicare Prescription Drug Plan	920058	963478	945794	893609	771694	815223	826907	1122209	991457
Medicare-Medicaid Plan HMO/HMOPOS	NA	NA	NA	NA	NA	NA	265	1319	4130
MN Disability Health Options	21	NA	NA	NA	NA	NA	NA	NA	NA
MN Senior Health Options	968	NA	NA	NA	NA	NA	NA	NA	NA
MSA	4422	16515	12267	135	6421	6416	6431	6449	6518
MSA Demo	3274	NA	NA	NA	NA	NA	NA	NA	NA
National PACE	405	548	616	717	781	858	953	1118	1216
PFFS	364285	630756	683361	385733	45781	36423	31919	24905	13658
Pilot	15	12	201	53	3	3	2	2	2
PSO (Federal Waiver of State License)	162	NA	NA	NA	NA	NA	NA	NA	NA
PSO (State License)	421	535	87	123	176	171	NA	NA	NA
Regional PPO	26402	27990	25943	24442	22773	21602	19970	19773	17578
RFB PFFS	NA	NA	3006	NA	NA	NA	NA	NA	NA
SHMO	1125	NA	NA	NA	NA	NA	NA	NA	NA
WI Partnership Program	42	NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	27505	277533	NA	NA	NA	NA	NA

Question 4

```
#filter for snp and for 800 series plan
# full_ma_data2 <- full_ma_data %>%
#                   filter(snp == 'No' & eghp == 'No'& !(planid %in% 800:899))
```

```
# q_4_df <- full_ma_data2 %>%
#   group_by(year, plan_type)%>%
#   summarize(n_under_plan_type = n()) %>%
#   spread(year, n_under_plan_type)

read_rds("results/q_4_df.rds")
```

plan_type	2007	2008	2009	2010	2011	2012	2013	2014	2015
1876 Cost	5855	5459	5825	6035	6851	7633	7731	7069	7157
Continuing Care Retirement Community	95	122	158	142	NA	NA	NA	NA	NA
Employer Direct PFFS	3247	NA	NA	NA	NA	NA	NA	NA	NA
Employer/Union Only Direct Contract PDP	32358	29113	25860	28700	28697	28669	25526	25528	25630
Employer/Union Only Direct Contract PFFS	NA	3332	3335	3332	3329	3323	NA	NA	NA
ESRD I	75	122	123	117	NA	NA	NA	NA	NA
ESRD II	12	12	7	8	NA	NA	NA	NA	NA
HCPP - 1833 Cost	13	13	3938	3604	11	11	10	9	9
HMO/HMOPOS	59948	70112	479962	506792	528465	507264	530900	523295	479266
Local PPO	17427	38470	405197	417551	515700	636701	633884	664716	704993
MA Health Senior Care Options	73	NA	NA	NA	NA	NA	NA	NA	NA
Medicare Prescription Drug Plan	920058	963478	945794	893609	771694	815223	826907	1122209	991457
Medicare-Medicaid Plan HMO/HMOPOS	NA	NA	NA	NA	NA	NA	265	1319	4130
MN Disability Health Options	21	NA	NA	NA	NA	NA	NA	NA	NA
MN Senior Health Options	968	NA	NA	NA	NA	NA	NA	NA	NA
MSA	4422	16515	12267	135	6421	6416	6431	6449	6518
MSA Demo	3274	NA	NA	NA	NA	NA	NA	NA	NA
National PACE	405	548	616	717	781	858	953	1118	1216
PFFS	364285	630756	683361	385733	45781	36423	31919	24905	13658
Pilot	15	12	201	53	3	3	2	2	2
PSO (Federal Waiver of State License)	162	NA	NA	NA	NA	NA	NA	NA	NA
PSO (State License)	421	535	87	123	176	171	NA	NA	NA
Regional PPO	25104	26692	25943	24442	22773	21602	19970	19773	17578
RFB PFFS	NA	NA	3006	NA	NA	NA	NA	NA	NA
SHMO	1125	NA	NA	NA	NA	NA	NA	NA	NA
WI Partnership Program	42	NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	27505	277533	NA	NA	NA	NA	NA

Question 5

```
# contract_service_area<- read_rds("data/output/contract_service_area.rds")
```

```
#Should I work with the filtered data here?
# joined_df <- left_join(full_ma_data2, contract_service_area ,by = c("contractid", "fips", "year"))
```

```
#What is meant by restrict dataset ot contracts that are approved by respective counties? Given the lef
# joined_df %>%
#   filter(!is.na(avg_enrollment))%>%
#   group_by(year, county.x)%>%
#   summarize(avg_prem = mean(avg_enrollment, na.rm = TRUE))%>%
#   group_by(year)%>%
#   summarize(avg = mean(avg_prem))%>%
#   ggplot(aes( x = year, y  = avg))+
```

```
#   geom_line()+
#   labs( title = 'Average of enrollees per county', x = 'Year', y = 'Number of enrollees')+
#   theme_minimal()
#How is approved by respective counties mean? What does non-missing enrollment data mean?
```

```
# library(png)
# readPNG("results/Q5_Graph.png")
```

Tried to read this in but it wouldn't work.

Question 6

```
# penetration_data <- read_rds("data/output/ma_penetration.rds")
# premium_data <- read_rds ("data/output/plan_premiums.rds")
```

```
# joined_df2 <- left_join(premium_data, penetration_data, by = c("state", "county", "year"))
```

```
# joined_df3 <- left_join(joined_df, joined_df2, by = c("fips",  "year", "contractid", "county.x" = "co
```

Joined to the wrong thing and it would give me an error and consequently I couldn't run the rest

```
# joined_df3 %>%
#   group_by(year)%>%
#   summarize(prem = mean(premium, na.rm = TRUE))%>%
#   ggplot( aes(year, prem))+
#   geom_line(color = 'red')+
#   labs( title = 'Average premium over time', x = 'Year', y = 'Average Premium')+
#   theme_minimal()
```

Question 7

```
# joined_df3 %>%
#   filter(!is.na(premium))%>%
#   group_by(year)%>%
#   summarize(perc_0 = ((sum(premium == 0))/n())* 100)%>%
#   ggplot( aes(year, perc_0))+
#   geom_line(color = 'blue')+
#   labs( title = 'Percentage of $0 Premium Plans over time', x = 'Year', y = 'Percentage of $0 Premium
#   theme_minimal()
```

Question 8

Question 9

Question 10

Working with this data was at times a little frustrating, since I was not really familiar with the data, and what the end goal was. I was not really working and doing all these things with a goal in mind, but instead I was simply following instructions, which made it harder for me to understand the purpose of the assingment and what I was supposed to achieve with it, and consequently at times I didn't know what I was doing.