Asthma Analysis

Sutianyi Wen

Date

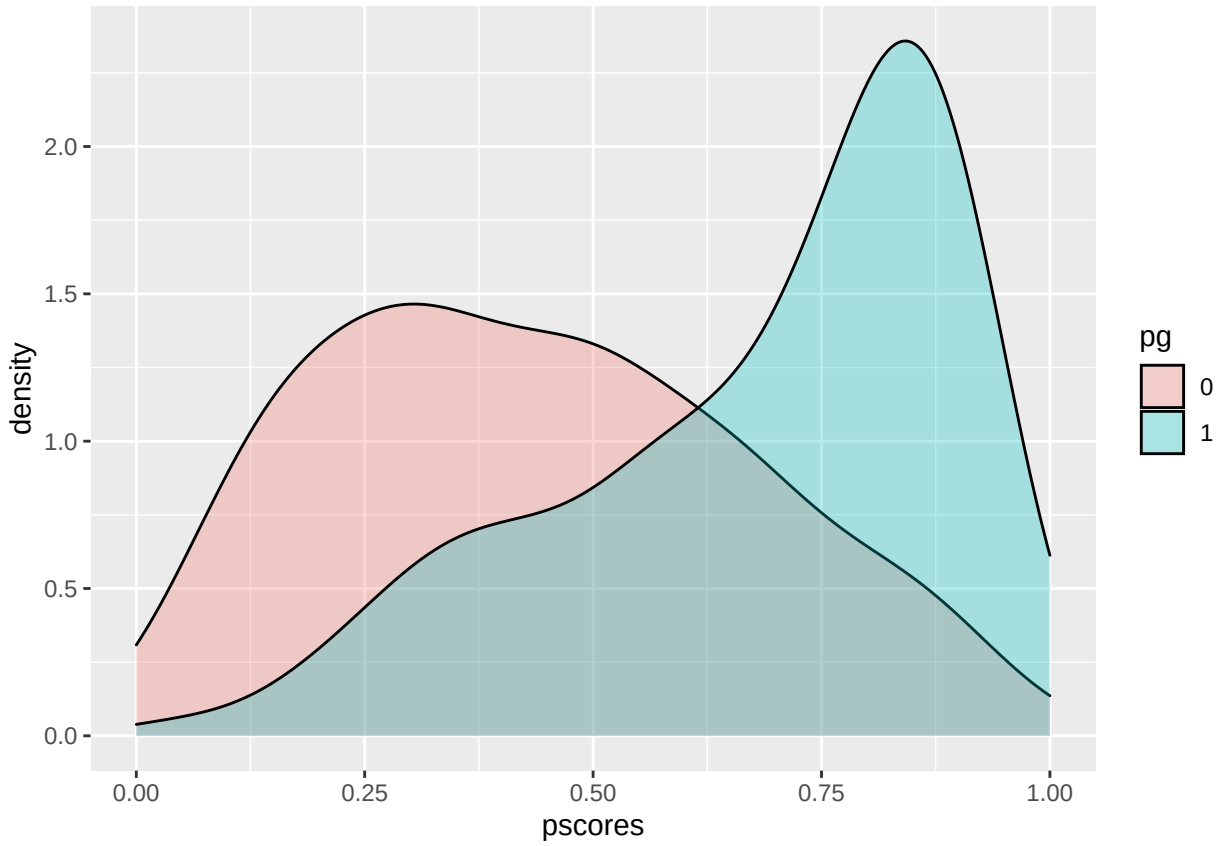
Question 1:

Unbalanced covariates are i_sex, i_race_1, i_race_2, i_educ_4, i_educ_5, com_t and pcs_sd. I use bal.tab to print mean differences for all predictors and levels. Predictors/levels have absolute value of mean difference greater than 0.1 are considered to be unbalanced.

	Type	Diff.Un
i_age	Contin.	-0.0940
i_sex	Binary	-0.1087
i_race_0	Binary	0.0292
i_race_1	Binary	-0.1965
i_race_2	Binary	0.2101
i_race_3	Binary	-0.0456
i_race_4	Binary	0.0027
i_educ_5	Binary	0.1706
i_educ_1	Binary	-0.0095
i_educ_2	Binary	-0.0075
i_educ_3	Binary	-0.0663
i_educ_4	Binary	-0.1650
i_educ_6	Binary	0.0778
i_insu_1	Binary	-0.0383
i_insu_2	Binary	0.0492
i_insu_3	Binary	0.0058
i_insu_5	Binary	-0.0167
i_drug	Binary	-0.0136
i_seve_3	Binary	-0.0158
i_seve_1	Binary	0.0802
i_seve_2	Binary	-0.0069
i_seve_4	Binary	-0.0575
com_t	Contin.	-0.9872
pcs_sd	Contin.	0.7537
mcs_sd	Contin.	-0.0531

Question 2(a):

According to the rang of propensity scores for two groups, there exist 8 outliers on the left tail and 40 outliers on the right tail. Therefore, we need to drop those observations to ensure the overlap.



Question 2(b):

After trying to identify potential interactions between variables, there are actually no interesting interactions can improve the quality of the matched dataset so I will stick to the original formula which includes all the variables.

After dropping 48 outliers from original dataset, covariates: i_sex, i_educ_5, i_educ_4, i_educ_6, com_t, pcs_sd and mcs_sd are still have Diff.Unit greater than 1, which means they are unbalanced. And most of the mean difference from summary are negative, which means the matched data has a even worse covariates balance. However, there is nothing we can do to improve it so we will go with what we had.

Question 2(c):

The causal effect Q is -0.20, which means patients' average satisfaction scores for physician group 1 is 0.20 lower than those for physician group 0 and the standard error is 0.07. The confidence interval for the causal effects are between -0.32 and 0.13, which means we are 95% confident that patients' average satisfaction scores for physician group 1 can be 0.32 lower than those for physician group 0 but also can be also 0.13 higher than those for physician group 0. And because the 95% confidence interval contains 0, we can't conclude there is a difference in patients' satisfaction scores between physical group 1 and 0.

	Value
Q	-0.20
Standard Error	0.07
Conf.Intv.Lower_bound	-0.32
Conf.Intv.Upper_bound	0.13

Question 2(d):

The estimated causal odds ratio is 0.32. Therefore, if a patient is from physician group 1 and everything else staying the same, the odds of being satisfactory is decreased by 68%. pg is significant.

Predictors	Odds Ratios	CI	p
(Intercept)	0.00	0.00 – 14.94	0.129
i_age	1.03	0.97 – 1.10	0.334
i_sex [1]	1.08	0.48 – 2.45	0.849
i_race [1]	9.61	0.32 – 291.74	0.194
i_race [2]	0.00	0.00 – Inf	0.994
i_race [3]	731240142.38	0.00 – Inf	0.991
i_race [4]	7.22	0.60 – 86.84	0.119
i_educ [2]	12441481.44	0.00 – Inf	0.993
i_educ [3]	0.67	0.04 – 11.76	0.784
i_educ [4]	1.27	0.14 – 11.50	0.830
i_educ [6]	0.51	0.21 – 1.25	0.141
i_insu [2]	1.25	0.30 – 5.31	0.760
i_insu [5]	2.63	0.26 – 26.42	0.411
i_drug [1]	41.69	0.56 – 3129.65	0.090
i_seve [1]	2.45	0.61 – 9.87	0.208
i_seve [2]	1.12	0.46 – 2.75	0.806
i_seve [4]	0.36	0.07 – 1.87	0.223
com_t	2.18	0.53 – 9.01	0.283
pcs_sd	0.90	0.78 – 1.04	0.168
mcs_sd	0.99	0.93 – 1.05	0.705
pg [1]	0.32	0.10 – 0.99	0.047
distance	553.72	0.00 – 164221158.76	0.326

Question 2(e):

I used one-to-n matching here and there are still unbalanced covariates: i_educ_5, i_educ_4, com_t and pcs_sd. According to summary on matched, although there are several negative values such as i_age meaning the balance gets worse from matching, most of the covariates' Diff.Unit is till less than 0.1 so we don't have to worry about them.

The causal effect Q is -0.15, which means patients' average satisfaction scores for physician group 1 is 0.15 lower than those for physician group 0 and the standard error is 0.06. The confidence interval for the causal effects are between -0.28 and 0.13, which means we are 95% confident that patients' average satisfaction scores for physician group 1 can be 0.28 lower than those for physician group 0 but also can be also 0.13 higher than those for physician group 0. And because the 95% confidence interval contains 0, we can't conclude there is a difference in patients' satisfaction scores between physical group 1 and 0.

	Value
Q	-0.15
Standard Error	0.06
Conf.Intv.Lower_bound	-0.28
Conf.Intv.Upper_bound	0.13

The estimated causal odds ratio is 0.46. Therefore, if a patient is from physician group 1 and everything else staying the same, the odds of being satisfactory is decreased by 54%. pg is significant.

Predictors	Odds Ratios	CI	p
(Intercept)	0.00	0.00 – 27.59	0.181
i_age	1.02	0.97 – 1.09	0.406
i_sex [1]	1.23	0.60 – 2.55	0.573
i_race [1]	5.99	0.29 – 124.45	0.248
i_race [2]	0.00	0.00 – Inf	0.994
i_race [3]	226103326.88	0.00 – Inf	0.991
i_race [4]	4.42	0.48 – 40.61	0.189
i_educ [2]	16684901.47	0.00 – Inf	0.993
i_educ [3]	0.67	0.06 – 7.63	0.746
i_educ [4]	1.00	0.14 – 7.09	0.999
i_educ [6]	0.45	0.20 – 1.05	0.065
i_insu [2]	0.93	0.25 – 3.49	0.917
i_insu [5]	1.63	0.19 – 13.57	0.653
i_drug [1]	23.07	0.42 – 1269.75	0.125
i_seve [1]	2.24	0.63 – 7.92	0.211
i_seve [2]	1.22	0.55 – 2.72	0.628
i_seve [4]	0.43	0.10 – 1.78	0.244
com_t	1.85	0.53 – 6.46	0.337
pcs_sd	0.92	0.80 – 1.04	0.179
mcs_sd	0.99	0.93 – 1.04	0.616
pg [1]	0.46	0.22 – 0.99	0.048
distance	115.09	0.00 – 5028684.35	0.384

The results of using a one-to-n matching are the same the one-to-one matching method. Their causal effects both have a 95% confidence interval that contains 0 and pg is significant in both models.

Question 2(f):

I think **one-to-n** matching works better in this case because it has less unbalanced covariates compared to one-to-one matching in our example. Besides matching methodology, it's more appropriate to use **regression** to measure the causal effects compatared to computer the average causal effect Q directly because there are unbalanced covariates in the data set.
