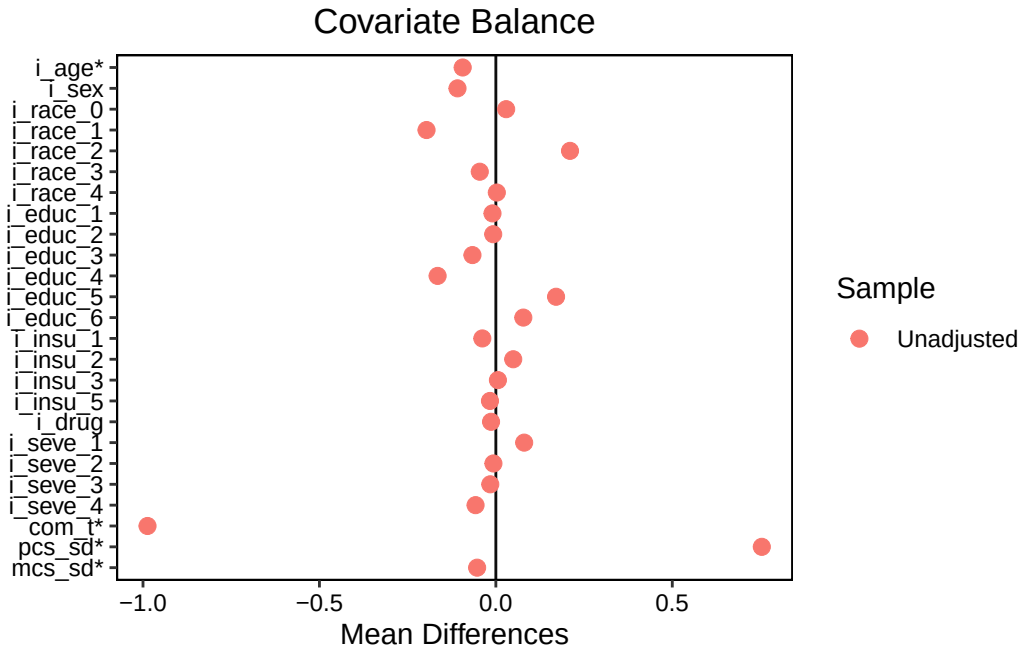
702 Assignment V

Question 1

First, centered `com_t`, `pcs_sd`, and `mcs_sd`, and changed the scale of `pg` from 1-2 to 0-1. After that, converted `pg` to categorical type.

Using bal.tab and love plot, it is noticeable that `com_t`, `pcs_sd`, `i_race`, and `i_educ` have high mean difference across physician group 0 and 1 (|mean|>0.1). Thus, it is evident that our covariates are not balanced across the two groups. Besides, treatment group has a sample size of 173 while control group has 105. In order to implement one-to-one matching, the inidicator for treatment and control group are switched such that physician group 2 becomes the control group and group 1 becomes the treatment group. This information is stored in the new indicator variable `treatment_s`.



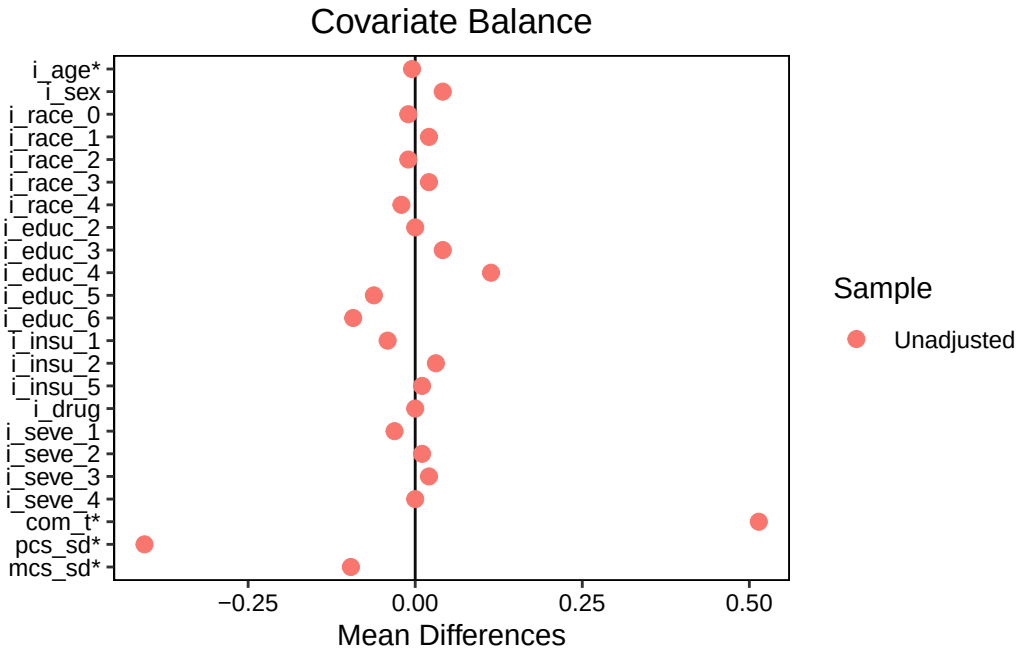
Question 2

a

After fitting a logistic regression with all pre-treatment variables as main effects to obtain propensity score `e`, we find 48 outliers. In order to ensure proper overlap, we choose to discard all 48 observations.

b

Using one-to-one, nearest neighbor matching on the estimated propensity scores to match observations from treatment group and control group in order to balance covariates. After checking balance again, we find that most of covariates have mean difference less than 0.1, except `i_educ_4`, `com_t`, and `pcs_sd`.



c

After calculating `p1 - p2` using matched data, we know that the average causal effect is 0.155, with a standard error of 0.065. This means that the difference in the proportion of patients being satisfied in the treatment group (physician group 1) is 15.5% higher than that in the control group (physician group 2). Based the results, our 95% confidence interval for the average causal effect is from 2.8% to 28.2%. This means that we are 95% confident that the true difference of proportion of patients being satisfied in the treatment versus control physician group is between 2.8% and 28.2%. Since this interval does not include 0, we can reject the null hypothesis.

d

The logistic regression model shows that the coefficient for `treatment_s` (the indicator variable for treatment and control group) is 0.92033, meaning that the causal odds ratio of having satisfaction = 1 being in the treatment group is 1.51 times than that of the control group, holding other variables constant.

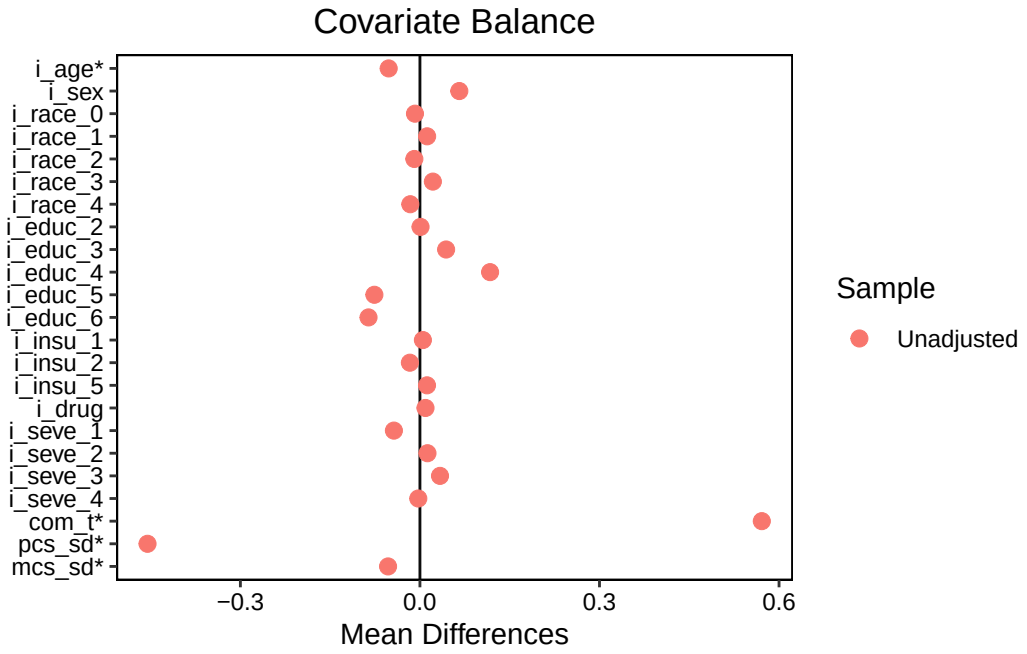
term	estimate	std.error	statistic	p.value
(Intercept)	34.6183877	2694.9219556	0.0128458	0.9897508
i_age	0.0303777	0.0392077	0.7747902	0.4384636
i_sex1	0.5023647	0.4495348	1.1175211	0.2637716
i_race1	1.0017694	2.0666714	0.4847261	0.6278707
i_race2	-17.7875295	2197.7341520	-0.0080936	0.9935423
i_race3	17.1745196	1419.0813789	0.0121026	0.9903438
i_race4	0.8837998	1.4489968	0.6099391	0.5419022
i_educ3	-17.2003418	1714.9351911	-0.0100297	0.9919976
i_educ4	-17.2354969	1714.9351803	-0.0100502	0.9919812
i_educ5	-16.5693680	1714.9352874	-0.0096618	0.9922911
i_educ6	-17.5364056	1714.9353398	-0.0102257	0.9918412
i_insu2	-0.0120868	0.9589336	-0.0126044	0.9899434
i_insu5	-0.7302791	1.1873891	-0.6150294	0.5385353
i_drug1	-16.6977494	2078.8461117	-0.0080322	0.9935913
i_seve2	-1.2608783	0.9005368	-1.4001407	0.1614712

term	estimate	std.error	statistic	p.value
i_seve3	-1.9222638	0.9276072	-2.0722822	0.0382391
i_seve4	-2.5887862	0.9090381	-2.8478302	0.0044018
mcs_sd	-0.0351567	0.0414978	-0.8471940	0.3968870
com_t	0.3227999	0.8861178	0.3642856	0.7156448
pcs_sd	-0.0750555	0.0903656	-0.8305757	0.4062134
distance	-2.5768066	7.2636596	-0.3547532	0.7227745
treatment_s1	0.9203286	0.4123543	2.2318878	0.0256224

e

e. b

Using one-to-five, nearest-neighbor matching, we also find imbalance of covariates given that `i_educ 4`, `com_t`, `pcs_sd` with mean differences larger than 0.1



e.c

After calculating `p1 - p2` using one-to-five, with-replacement matched data, we know that the average causal effect is 0.166, with a standard error of 0.064. This means that the difference in proportion of patients being satisfied in the treatment group (physician group 1) is 16.6% higher than that in the control group (physician group 2). Based the results, our 95% confidence interval for the average causal effect is from 2.9% to 28%. This means that we are 95% confident that the true difference of proportion of patients being satisfied in the treatment versus control physician group is between 2.9% and 28%. Since this interval does not include 0, we can reject the null hypothesis. Although there is slight difference in number, the conclusion of average causal effect drawn from this data is the same with that from the one-to-one, with-replacement data.

e.d

The logistic regression model on one-to-five, with-replacement matched data shows that the coefficient for `treatment_s` (the indicator variable for treatment and control group) is 0.8779, meaning that the causal

odds ratio of having satisfaction = 1 being in the treatment group (physician group 1) is 1.405 times hgiher than that of the control group (physician group 2), holding other variables constant.

Compared to the 1.51 causal odds ratio obtained from the one-to-one, without-replacement data, the new matched data offers a smaller causal effect estimate.

term	estimate	std.error	statistic	p.value
(Intercept)	16.9810593	1679.1407981	0.0101129	0.9919312
i_age	0.0240186	0.0324859	0.7393537	0.4596923
i_sex1	0.0300828	0.4252662	0.0707389	0.9436056
i_race1	1.5161378	1.7599506	0.8614661	0.3889814
i_race2	-18.1346144	2219.4625288	-0.0081707	0.9934808
i_race3	18.6325661	1528.9926991	0.0121862	0.9902771
i_race4	1.4278900	1.3172864	1.0839632	0.2783811
i_educ3	-16.9140674	1679.1390416	-0.0100731	0.9919630
i_educ4	-16.6052251	1679.1389684	-0.0098891	0.9921097
i_educ5	-16.3139251	1679.1390365	-0.0097156	0.9922482
i_educ6	-17.5665966	1679.1390651	-0.0104617	0.9916529
i_insu2	-0.3070045	0.7811750	-0.3930035	0.6943169
i_insu5	-0.0374103	1.0790215	-0.0346706	0.9723424
i_drug1	0.8249999	2.0430432	0.4038093	0.6863529
i_seve2	-0.4831266	0.7183536	-0.6725472	0.5012354
i_seve3	-0.8917323	0.7102169	-1.2555775	0.2092692
i_seve4	-1.5918513	0.8079931	-1.9701299	0.0488235
mcs_sd	-0.0188206	0.0333276	-0.5647149	0.5722677
com_t	0.6314667	0.7147815	0.8834402	0.3769985
pcs_sd	-0.1029844	0.0742555	-1.3868936	0.1654742
distance	-4.7118236	6.0224393	-0.7823779	0.4339925
treatment_s1	0.8779076	0.4008318	2.1902148	0.0285087

Question 3

Between using one-to-one, without-replacement or one-to-many, with-replacement method for estimating causal effects, I feel more comfortable using the former. This is because that allowing one-to-many and replacement will inevitably lead to an over-representation of certain control-group observations, since given their attributes they can be matched to different treatment observations. Comparing calculating average causal effect using mean of proportion and logistic regression, I prefer the latter given that it provides an estimate while controlling for other covariates.