Section 3 Assignment

Question 1: Why do the logistic, log, and identity link functions enable us to interpret coefficient as a log odds ratios, log risk ratios, and risk differences? Please show the math. Below are the equations that prove how the logistic, log, and identity link functions enable us to interpret coefficient as a log odds ratios, log risk ratios, and risk differences.

Log Odds Ratio General Form:

$$\log[\frac{P(Y=1|X)}{1-P(Y=1|X)}] = \beta_0 + \beta_1 X$$

When X=1

$$\log[(\frac{P(Y=1|X=1)}{1-P(Y=1|X=1)}] = \beta_0 + \beta_1 * 1 = \beta_0 + \beta_1$$

When X=0

$$\log[\frac{P(Y=1|X=0)}{1-P(Y=1|X=0)}] = \beta_0 + \beta_1 * 0 = \beta_0$$

$$\log[\frac{\frac{P(Y=1|X=1)}{1-P(Y=1|X=1)}}{\frac{P(Y=1|X=0)}{1-P(Y=1|X=0)}}] = \beta_0 + \beta_1 - \beta_0 = \beta_1$$

Log Risk Ratio General form:

$$\log[P(Y=1|X)] = \beta_0 + \beta_1 X$$

When X=1

$$\log[P(Y=1|X=1)] = \beta_0 + \beta_1 * 1 = \beta_0 + \beta_1$$

When X=0

$$\log[P(Y=1|X=0)] = \beta_0 + \beta_1 * 0 = \beta_0$$

$$\log[\frac{P(Y=1|X=1)}{(1-P(Y=1|X=0))}] = \beta_0 + \beta_1 - \beta_0 = \beta_1$$

Log risk difference-

General Form:

$$\log[P(Y=1|X)] = \beta_0 + \beta_1 X$$

When X=1

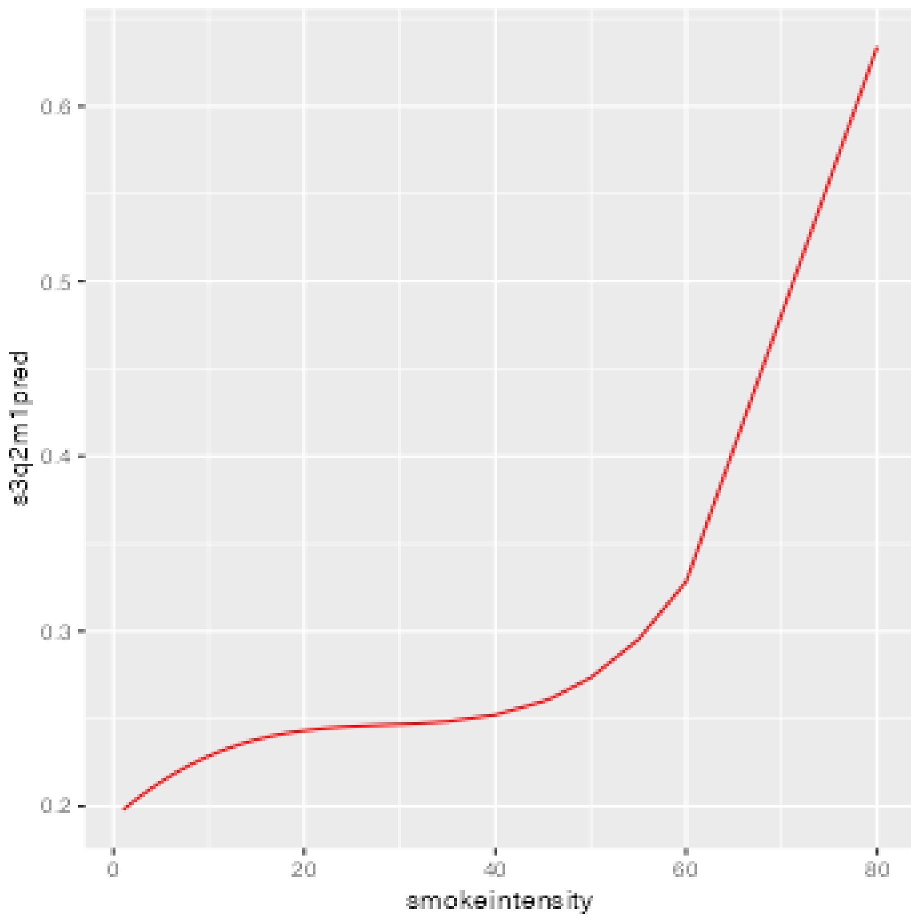
$$\log[P(Y=1|X=1)] = \beta_0 + \beta_1 * 1 = \beta_0 + \beta_1$$

When X=0

$$\log[P(Y=1|X=0)] = \beta_0 + \beta_1 * 0 = \beta_0$$

$$\log[P(Y=1|X=1)] - (1 - \log[P(Y=1|X=0)]) = \beta_0 + \beta_1 - \beta_0 = \beta_1$$

Question 2) Using the NHEFS dataset (available on CANVAS), please plot the unadjusted dose response relation between smoking intensity and the *risk* of high blood pressure.



Question 3) Using the NHEFS data and an outcome regression model, estimate the conditionally and marginally adjusted risk ratio and risk difference for the association between quitting smoking and high blood pressure. Adjust for smokeintensity, sex, age, race, school, and marital status. Please use appropriate coding for all variables, but do not adjust for interaction effects. As always, use appropriate standard error estimators. %>% select(qsmk, high_BP, smokeintensity, sex, age,race,school,marital)

The marginally adjusted risk ratio is 0.537 (0.404, 0.667) The marginally adjusted risk difference is -0.124 (-0.164, -0.084)

#Conditional Portion

The conditional RR is 0.540 SE: 0.119 The conditional RD is -0.127 SE: 0.021

Question 4) For the risk difference model used in question 3, are there any relevant interaction effects?

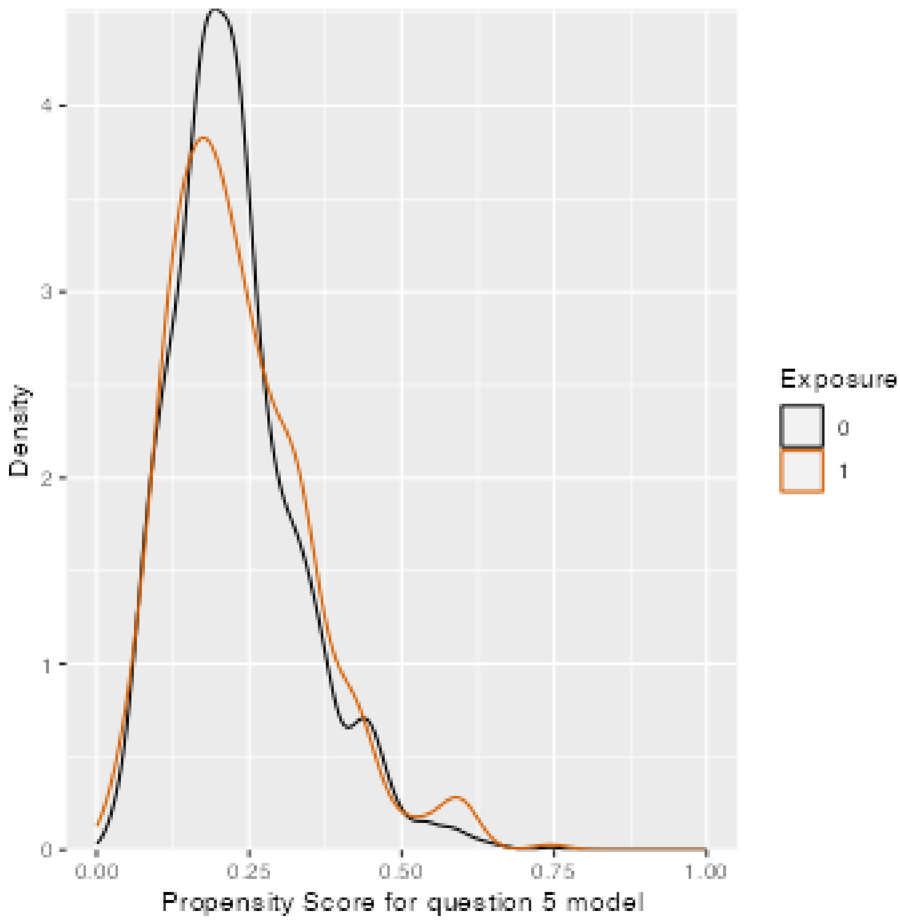
The significant interactions for the data for the model from question 3 are:

- qsmk:sex P = 0.01173 *
- qsmk:race P = 0.00590 **
- smokeintensity:race P = 0.00465 **
- sex:race P = 0.03608 *
- sex:marital P = 0.01915 *
- age:race P = 0.00441 **
- school:marital P = 0.03038 *

Question 5: Using the NHEFS data and a propensity score model with IP weighting, estimate the marginally adjusted risk ratio and risk difference for the association between quitting smoking and high blood pressure. Adjust for smokeintensity, sex, age, race, school, and marital status. Please use appropriate coding for all variables, but do not adjust for interaction effects. As always, use appropriate standard error estimators.

Using propensity score model with IP weighting, the marginally adjusted risk difference: -0.116 SE: 0.012 Using propensity score model with IP weighting, the marginally adjusted risk ratio: 0.573 SE: 0.131 ***

Question 6: Is the positivity assumption met for the propensity score model fit in Question 5? Is the positivity assumption required to interpret the estimate in Question 5 AND Question 3 as a causal contrast of potential outcomes? Why?



When looking at the graph, positivity is not violated for question 6 model. This is because the mass of density for the exposed (orange) occurs in essentially the same place as the density mass for the unexposed (black).

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.5214 0.8209 0.9507 1.0003 1.1231 2.7366

The mean of the stabilized weights is essentially 1. The max weight is not large relative to the mean and min. This suggests that the weights are "well behaved." Thus, in this particular case, we are not concerned with violations of the positivity assumption.

Positivity assumption is required to interpret the estimate in Question 3 and 5 because we are using the average treatment effect. We need to ensure that there are exposed and unexposed individuals on all levels in order to ensure that the positivity assumption is not violated.