PS C236A / Stat C239A
Problem Set 3 - Solutions

1: First we need to show that $Z \perp X|e(X)$, which is equivalent to showing $P(Z|X) = P(Z|e(X))$.

$P(Z|e(X)) = E[Z|e(X)] = E[E[Z|e(x), x]|e(x)] = E[P(Z|X, e(X))|e(x)] = E[e(x)|e(x)] = e(x) = P(Z|X)$

Now we need to show that conditioning on the propensity score and under the stated assumptions, that the ATT is identified. We want to estimate $E((r_1 - r_0)|Z = 1)$ which is the ATT estimand. Note that conditioning on $Z = 1$ is equivalent to conditioning on $X|Z = 1$, which–as we proved above–is equivalent to conditioning on $e(X)|Z = 1$.

We can observe without assumptions: $E(r_1|Z = 1) - E(r_0|Z = 0)$ Because we assume that $r_0 \perp Z$, then $E(r_0|Z = 0)$ can be rewritten as $E(r_0|Z = 1)$. As a result, $E(r_1|Z = 1) - E(r_0|Z = 1) = E(r_1 - r_2|Z = 1)$

2) a) Note that these p_i we want to estimate are propensity scores. From Rosenbaum and Rubin, the probability of observing a given treatment assignment is

$$\prod_{i=1}^{10000} e(X_i)^{T_i}(1 - e(X_i))^{T_i}$$

where $e(X_i)$ is the propensity score given observed covariates X_i. From our problem statement, we know that $e(X_i)$ is a function of sex, exercising more than 30 minutes a day, and watching TV for more than an hour a day. Let S_i, E_i, and V_i denote indicator variables for these covariates. It follows that the probability of our treatment assignment is:

$$\prod_{i=1}^{10000} e(X_i)^{T_i}(1 - e(X_i))^{1-T_i} = \prod_{i=1}^{10000} e(S_i, E_i, V_i)^{T_i}(1 - e(S_i, E_i, V_i))^{1-T_i}$$
$$= \prod_{(s,e,v)\in\{0,1\}^3} e(s, e, v)^{\#(s,e,v,t)}(1 - e(s, e, v))^{\#(s,e,v)-\#(s,e,v,t)} \quad (1)$$

Here, s, e, and v, all take values equal to 0 or 1; $\#(s, e, v)$ denotes the number of units that have covariate indicators $S_i = s$, $E_i = e$, and $V_i = v$; and $\#(s, e, v, t)$ denotes the number of treated units that have those values of indicator variables.

We take the log of (**??**) to obtain:

$$\sum_{(s,e,v)\in\{0,1\}^3} \#(s, e, v, t) \log(e(s, e, v)) + (\#(s, e, v) - \#(s, e, v, t)) \log(1 - e(s, e, v))$$

We choose one of the eight possible choices of (s, e, v), and denote this choice as (s^*, e^*, v^*) Taking the