PS 6

a) Yes, we can look at the effect of gender on the likelihood of being admitted to college.

ie $\Pr(y=1) = \alpha_{1i} + \beta_{1i} G_i + \varepsilon_i$

b) independence btw. treatment & potential outcomes

ie

$$E(y_{i0} | D_i) = E(y_{i1} | D_i = 0) = E(y_{i1} | D_i = 1)$$

$$\Rightarrow E[y_i | D_i = 1] - E[y_i | D_i = 0] = \underbrace{E[y_{i1} | D_i = 1] - E[y_{i0} | D_i = 1]}_{ATT} + \underbrace{E(y_{0i} | D_i = 1) - E(y_{0i} | D_i = 0)}_{\text{selection}}$$

what would be outcome if they were not female

female

male

$$ATE = E(y_{i1}) - E(y_{i0}) = p_T \cdot (ATT) + (1 - p_T) ATU$$

Here under unconditional independence we just have $E(\text{admission}|\text{men}) - E(\text{admission}|\text{woman}) = \underline{\underline{14\%}}$ → calculated from the excel file attached

c) Even though gender assignment is exogenous (assuming no-gender specific abortions & that the characteristics of the parents who decided to take this path would influence admissions status), there is supposedly some selection into different study areas based on gender, so we can't talk about full independence

average within groups
& take average of these averages

d) conditional randomisation ATE

ATE : -4,26%

ATT : - 7,10 %

e) Generally they are not the same because there are diff. numbers of people in each department/field of study, ie, the weights we are using are different.

f) see do file. Coefficient of -0,018.

↳

g) OLS & matching not the same
- different weights are used
→ OLS is a linear matching estimator & is parametric, matching is non-parametric
→ OLS uses the full set of observations, matching only uses the matched observations

$\boxed{6.2}$

For B_2 to be consistent we need:

$cov(educ_i, u_i | X_i) = 0$

→ it's likely that some unobservables influencing education & health are the same → ie

$X)$ $b^{OLS} \xrightarrow{p} B + \frac{cov(ed, u)}{Var(ed)}$ $\neq 0$

b) Yes, if we can find a valid instrument (exogenous & relevant), then the problem in a can be solved.
This means we will only be using the variation coming from the instrument in order to estimate causal effects

c) Assumptions

ie no spillovers

① SUTVA: potential outcomes only depend on own instrument

educ(instrument) = educ(instrument$_i$) Health(educ, instrument) = health(educ$_i$, inst$_i$)

$h(M) = h(M_i)$ $\quad h(M, Z) = h(M_i, z_i)$

2) independence $\{h_i(M_{1i}, 1), h_i(M_{0i}, 0), M_{1i}, M_{0i}\} \perp\!\!\!\perp Z_i$

structural: exogeneity ie $cov(u, Z) \neq 0$

→ potential outcomes are independent of education

3) exclusion $h_i(M, 0) = h_i(M, 1)$ -

structural $cov(u, Z) = 0$

4) First stage $E(M_{1i} - M_{0i}) \neq 0$

structural $cov(Z, M) \neq 0$

5) Monotonicity $M_{1i} - M_{0i} \geq 0 \quad \forall_i$ → effects of education on health outcomes go into the same direction for anyone

d) compliers: $Z=1 \Rightarrow H=1$ & $Z=0 \Rightarrow H=0$
- always comply with the assigned treatment study
  - being born in the last quarter of the year & graduating high school
  - → being born in 1st quarter of the year & not graduating high school

always takers: graduating high school regardless of quarter of birth

never takers: not graduating from high-school regardless of when you were born

defiers: - graduating HS when you're born in Q1
- not graduating HS when born in Q4

I Assume there are no defiers.

e) BJB & BJ's claims:

1) quarter of birth is not a valid instrument because:

a) it is not exogenous, $cov(z, \epsilon) \neq 0$
- there are some unobservables determining both quarter of birth as well as whether you complete high school
  - you would expect correlation only for q. of birth and education only w people w. high school diploma or less, but you can see these relationships also for above high school lvl

* <u>exclusion</u> restriction doesn't hold, ie education is not the only channel through which q. of birth influences

→ BJB 97 says quarter of birth can be associated w a bunch of other factors that influence ~~education~~ earnings directly (parents' income, physical & mental health etc)

→ weak instruments:
association btw. q of birth & education attainment is weak, so ~~then~~ even a small direct association btw q. of birth & earnings can yield a huge bias

→ failure of exogeneity

$$\beta^{IV} = \beta + \frac{cov(Z, U)}{cov(X, Z)}$$

weak 1st stage makes bias worse

f) Theoretically yes.
But the geographical variation of where the law is implemented or not would have to be exogenous ie states where there is a lot of poverty ~~should~~ & low education shouldn't be more likely to have these laws in place for example.

↓ all the other assumptions from (c) would also need to be satisfied

g) people who are more educated learn about what makes ~~em~~ a person healthy at school & are therefore more likely to do these things which in turn make them more healthy

i) educ_i (years of completed) = α_i + quarter at birth$_i$ + $SMSA_i$ + married$_i$ + regions$_{1-8,i}$ + birth year + income$_i$

$\text{health}_i = A_i + \widehat{\text{educ}}_i + SMSA_i + \text{married}_i + \text{regions}_{1-8,i}$ + birth year$_i$ + income$_i$

j: see code
→ IV estimates are much bigger, almost 3X as big as OLS ~~[illegible]~~ estimates

k. I use "estat firststage" command.
My first stage is super weak in all cases.
↳ This is a huge issue because even for small violations of exclusion restriction (which are possible), I'd get huge bias.

l. Add here I'd expect my 2SLS estimate to be lower (in absolute terms) because (coeff on the) education is picking up some of the omitted variable.

m: 2SLS coefficient would also be higher due to heterogeneous treatment effects
→ this instrument captures return for compliers (as long as the assumptions of the LATE theorem hold]
↳ given this, it's possible that returns to education to health are larger than 2SLS.

n) i: No, I don't think our instruments are valid here. Weakness is a particular problem