a

Interpret all parameters (p,theta0, theta1, beta0, beta1, beta2).

*p: probability of mice obese at baseline

*theta0: The probability of mice get exposed to light without baseline covariate

*theta1: The probability of obese mice get exposed to light is 1/5 times lower than not obese mice

*beta0: Mean of glucose at baseline don't get exposed to light variable and obesity covariate, is 110

*beta1: Estimated coefficients of the obesity, glucose increase 20 units increase 20 units if obesity increase one unit

*beta2: Estimated coefficients of the light variable, glucose decrease 5 units if exposed to light versus do not exposed. (But in the code, beta2 equal 5 so glucose will increase 5 units if exposed to light versus do not exposed)

b

Write the marginal and conditional PACE. Under which assumptions marginal and conditional PACE are identify?

$$E[Y \mid A, C] = \beta_0 + \beta_1 \cdot C + \beta_2 \cdot A = 110 + 20C - 5A$$

Marginal PACE

$$\begin{aligned} E[Y_{1i} - Y_{0i}] &= (\beta_0 + \beta_1 + \beta_2)\frac{1}{5} + (\beta_0 + \beta_2)(1 - \frac{1}{5}) \\ &- (\beta_0 + \beta_1)\frac{1}{5} - \beta_0(1 - \frac{1}{5}) \\ &= 2\beta_2 \end{aligned}$$

conditional PACE

$$\begin{aligned} E[Y_{1i} \mid C] - E[Y_{0i} \mid C] &= \beta_0 + \beta_1 C + \beta_2 - \beta_0 + \beta_1 C \\ &= \beta_2 \end{aligned}$$

Under the assumptions of exchangeability marginal and conditional PACE are identify and so that exposure treatment are randomly assigned. Also we need SUTVA.

c

Show the g-formula for the randomized study we considered in the previous homework in which we studied the effect of light (dark=DL Vs bright=LL) on obesity. Compare the g-formula of the current simulated observational study.

*The g-formula of the previous homework is $E(Y_a) = E[Y|A = a]$

*Without considering the unobserved confounder, the g-formula of the this homework is
$E(Y_a) = \sum_c E[Y|A = a, C = c]P(C = c)$

*since in the previous homework, the exposure A(light) is independent with confounder, however in this homework, we need to consider the confounder since the probability of the treatment is different with different situation of the confounder C(obesity)

d

Provide the estimate of $E[Y \ |A = 1] \ - E[Y \ |A = 0]$ in your simulated data from part (a). Interpret the results.

```
> mean(Y[A==1])-mean(Y[A==0])
[1] -2.566766
```

The causal effect means glucose will decrease 2.57 when mice expose to variable light versus do not exposed to light when obesity status covariate stay constant.

e

Provide the estimate of $E[Y1] - E[Y0]$ in your simulated data from part (a) using the g-formula. Interpret the results. Explain the differences between the inferences obtained in (d) and (e).

If we don't use boostrap.

$E(Y_1) - E(Y_0) = \sum_c E[Y|A = 1, C = c]P(C = c) - \sum_c E[Y|A = 0, C = c]P(C = c)$=-22.02507

```
dt = data.table(Y=Y,A=A,C=C)

#without boostrap
mean(dt[A==1][C==0]$Y)
mean(dt[A==1][C==1]$Y) #means no data
mean(dt[A==0][C==0]$Y)
mean(dt[A==0][C==1]$Y)

E_Y1 = mean(dt[A==1][C==0]$Y)*0.8
E_Y0 = mean(dt[A==0][C==0]$Y)*0.8 + mean(dt[A==0][C==1]$Y)*0.2
E_Y1-E_Y0
```

If we use boostrap. This causal effect means glucose will increase 4.29 when mice expose to variable light versus do not exposed to light when obesity status covariate stay constant. Compare to d, we obtain positive causal effect and it is more understandable because in the code the beta is positive which means that glucose will increase if exposed to light versus do not exposed.

f

Consider now a hypothetical observational study with 10 continuous covariates. Under which assumptions can you estimate E[Y1] −E[Y0] using linear regression.

Follow lat of iterated expectation and randomization as well as consistency we could estimate this.