Denis Ostroushko - HW1

Problem 1

Table 1: All Potentially Confounding Variables Stratified by Treatment Group

	Control Group	Treatment Groud	SMD
N	406	184	
Pre-term Pregnancy = Yes (%)	53 (13.1)	18 (9.8)	0.103
Birth.outcome (%)			0.271
Elective abortion	1 (0.2)	1 (0.5)	
Live birth	391 (96.3)	183 (99.5)	
Non-live birth	14 (3.4)	0 (0.0)	
Clinic (%)			0.150
KY	103 (25.4)	45 (24.5)	
MN	123 (30.3)	58 (31.5)	
MS	96 (23.6)	52 (28.3)	
NY	84 (20.7)	29 (15.8)	
Age (mean (SD))	25.91 (5.52)	25.99 (5.60)	0.015
Black = Yes (%)	180 (44.3)	82 (44.6)	0.005
White = Yes (%)	118 (29.1)	27 (14.7)	0.354
Nat.Am = Yes (%)	117 (28.8)	80 (43.5)	0.309
Asian = Yes (%)	4 (1.0)	0 (0.0)	0.141
Hisp (%)			0.655
No	159 (39.2)	99 (53.8)	
Yes	178 (43.8)	85 (46.2)	
NA	69 (17.0)	0 (0.0)	
Education (%)			0.032
8-12 yrs	238 (58.6)	107 (58.2)	
LT 8 yrs	76 (18.7)	33 (17.9)	
MT 12 yrs	92 (22.7)	44 (23.9)	
Public Assistance = Yes (%)	309 (76.1)	136 (73.9)	0.051
Hypertension = Y (%)	9 (2.2)	6 (3.3)	0.064
Diabetes = Yes (%)	8 (2.0)	9 (4.9)	0.161
BL.Diab.Type (%)			0.191
Type I	1 (0.2)	4 (2.2)	
Type II	7 (1.7)	5 (2.7)	
NA	398 (98.0)	175 (95.1)	
BMI (mean (SD))	27.49 (6.89)	28.04 (7.95)	0.074
Use.Tob (%)			0.106
No	353 (86.9)	165 (89.7)	
Yes	44 (10.8)	17 (9.2)	

NA	9 (2.2)	2 (1.1)	
BL.Cig.Day (mean (SD))	0.82 (3.09)	0.90 (3.43)	0.025
Use.Alc (%)			0.090
No	389 (95.8)	178 (96.7)	
Yes	8 (2.0)	4 (2.2)	
NA	9 (2.2)	2 (1.1)	
BL.Drks.Day (mean (SD))	0.04 (0.40)	0.03 (0.38)	0.013
Drug.Add = NA (%)	10 (2.5)	2 (1.1)	0.104
Prev.preg = Yes (%)	303 (74.6)	133 (72.3)	0.053
N.prev.preg (mean (SD))	1.87 (1.82)	1.67 (1.73)	0.114
Live.PTB = Yes (%)	43 (10.6)	9 (4.9)	0.215
Any.stillbirth = Yes (%)	6 (1.5)	2 (1.1)	0.035
Spont.ab = Yes (%)	94 (23.2)	52 (28.3)	0.117
Induced.ab = Yes (%)	67 (16.5)	8 (4.3)	0.406
Any previous adverse birthing events = Yes (%)	169 (41.6)	62 (33.7)	0.164
N.living.kids (mean (SD))	1.87 (1.82)	1.67 (1.73)	0.114
N.qualifying.teeth (mean (SD))	14.33 (6.67)	14.41 (6.92)	0.011
BL.GE (mean (SD))	1.42 (0.39)	1.49 (0.42)	0.173
BL..BOP (mean (SD))	69.12 (17.04)	70.08 (16.99)	0.057
BL.PD.avg (mean (SD))	2.84 (0.53)	2.87 (0.59)	0.064
BL..PD.4 (mean (SD))	24.80 (15.94)	25.67 (15.48)	0.055
BL..PD.5 (mean (SD))	9.95 (13.41)	10.09 (13.98)	0.010
BL.CAL.avg (mean (SD))	1.37 (0.66)	1.50 (0.78)	0.176
BL..CAL.2 (mean (SD))	41.29 (25.27)	46.10 (27.47)	0.183
BL..CAL.3 (mean (SD))	14.03 (16.05)	16.91 (17.80)	0.170
BL.Calc.I (mean (SD))	1.13 (0.62)	1.17 (0.61)	0.066
BL.Pl.I (mean (SD))	1.23 (0.48)	1.28 (0.49)	0.111
GA.at.outcome (mean (SD))	269.18 (26.50)	272.29 (17.48)	0.139
Birthweight (mean (SD))	3180.82 (727.49)	3259.16 (574.17)	0.120
Preg.ended (mean (SD))	0.13 (0.34)	0.10 (0.30)	0.103

Problem 2

Unadjusted Treatment effect for reduction in pre-term pregnancies

- Unadjusted Average Treatment Effect (ATE) is the difference is proportions of pre-term pregnancies
- Control group experienced 13.1% of pre-term pregnancies
- Treatment group experienced 9.8% of pre-term pregnancies
- Average Treatment Effect is the difference between two means, which is -3.3%
- Variance of the difference for two independent random variables is the sum of their variances, which works out to be 7.59×10^{-4}. Therefore, standard error for the test is 0.0276
- In order to see if the the proportion decreased for the treatment group, compared with the control group, we perform a one sided z-test:
- 5% percentile of a standard normal is -1.65

- test statistic is $\frac{0.0327-0}{0.027} = -1.18$
- Since the test statistic is does not fall below the cutoff, we conclude that there are no statistically significant differences in the proportion of pre-term pregnancies between the control and treatment groups
- Conclusion:
  - Treatment effect: -0.033
  - Standard error 0.028
  - 95% Confidence interval: (-0.087, 0.021)

Unadjusted Treatment effect for increase in infant birthweights

- Unadjusted Average Treatment Effect (ATE) is the difference is average birthweights between the two groups
- Control group showed average birthweight of 3180.8
- Treatment group showed average birthweight of 3259.2
- Average Treatment Effect is the difference between two means, which is 78.3%
- Variance of the difference for two independent random variables is the sum of their variances, which works out to be 3095.21. Therefore, standard error for the test is 55.63
- In order to see if the the proportion decreased for the treatment group, compared with the control group, we perform a one sided z-test:
- 95% percentile of a standard normal is 1.65
- test statistic is $\frac{78.33922-55.63465}{0.027} = 1.408101$
- Since the test statistic is does not fall above the cutoff, we conclude that there are no statistically significant differences in the proportion of pre-term pregnancies between the control and treatment groups
- Conclusion:
  - Treatment effect: 78.339
  - Standard error 55.635
  - 95% Confidence interval: (-30.705, 187.383)

Problem 3

average (causal) treatment effect for pre-term pregnancy rates

In order to estimate potential outcomes Y^1 and Y^0 for each patient we fit the following regression model:

```
pregnancy_model <-
  glm(
    `Preg.ended...37.wk` ~
      Group + Race_ethnicity + Public.Asstce + Use.Tob + N.prev.preg +
      Live.PTB + BL.GE + BL..BOP + BL..PD.4 + BL..CAL.3,

    data = data,
    family = "binomial"
  )
```

Using this model, we obtain $E[Y^1] = 0.11$ and $E[Y^0] = 0.11$.

Therefore, the average causal treatment effect is 0

Using a bootstrap procedure with 5000 replications, we obtain a standard error for the average causal treatment effect of 0.03

Therefore, using regression models and bootstrap procedure we obtain:

- Average causal treatment effect: 0
- Standard error of the average causal treatment effect: 0.03
- 95% Confidence interval: (-0.05, 0.05)

average (causal) treatment effect for average birthweights

In order to estimate potential outcomes Y^1 and Y^0 for each patient we fit the following regression model:

```
borthweight_model <-
  lm(
    Birthweight ~
    Group *(Race_ethnicity + Public.Asstce + Use.Tob +
              N.prev.preg + Live.PTB + BL.GE + BL..BOP +
              BL..PD.4 + BL..CAL.3),
```

```
    data = data
  )
```

Using this model, we obtain $E[Y^1] = 3282.65$ and $E[Y^0] = 3225.43$.

Therefore, the average causal treatment effect is 57.22

Using a bootstrap procedure with 5000 replications, we obtain a standard error for the average causal treatment effect of 52.49

Therefore, using regression models and bootstrap procedure we obtain:

- Average causal treatment effect: 57.22
- Standard error of the average causal treatment effect: 52.49
- 95% Confidence interval: (-45.14, 159.58)