PHP 2610 HW 1

Blain Morin

December 12, 2018

Question 1:

For question 1, our goal is to fit the model:

$$logit(\pi(x)) = \beta_0 + \beta_1 brate + \beta_2 basewt + \beta_3 s12wgt + \epsilon$$

a.) Fit the model to complete cases only

We first fit a glm using the the complete cases only. Here is the summary of the model:

Table 1: Complete Cases Only

	Dependent variable:
	logit(Quit Status)
brate	-0.033
	(0.023)
basewt	-0.148^{***}
	(0.042)
s12wgt	0.157^{***}
	(0.041)
Constant	-1.727^{*}
	(1.006)
Observations	141
Log Likelihood	-81.530
Akaike Inf. Crit.	171.061
Note:	*p<0.1; **p<0.05; ***p<0.01

We see that while our data has 247 observations, only 141 are used in the complete case analysis.

b.) Use IPW for handling missingness in X3, and bootstrap resampling for standard error estimation.

Next, we use inverse probability weighting. Inverse probability weighting attempts to overcome the bias of using only complete cases by weighting the observed data by the inverse probability that they were observed. Here is a summary of the resulting model:

Table 2: IPW with Bootstrap

	Beta	SE	Z	p.value
Intercept	-2.266	1.263	-1.795	0.073
brate	-0.027	0.026	-1.036	0.300
basewt	-0.160	0.050	-3.198	0.001
s12wgt	0.167	0.047	3.533	0.0004

c.) Use either regression imputation with bootstrap, or multiple imputation, to handle missingness in X3.

Next we try imputing the missing data using the other variables in the data. Here are the results:

Table 3: Regression Imputation with Bootstrap Samples

	Beta	SE	Z	p.value
Intercept	-1.051	1.377	-0.764	0.445
brate	-0.036	0.023	-1.537	0.124
basewt	-0.295	0.086	-3.423	0.001
s12wgt	0.292	0.081	3.596	0.0003

Summary

Here we compare the beta coefficients for each of the models:

Table 4: Betas Coefficients From Each Model

	Complete.Cases	IPW	Reg.Imputation
(Intercept)	-1.727	-2.266	-1.051
brate	-0.033	-0.027	-0.036
basewt	-0.148	-0.160	-0.295
s12wgt	0.157	0.167	0.292

We see that the betas are similar between the complete cases model and the IPW model. The regression imputation model finds a larger effects for basewt and s12wgt.

Next, we compare the p-values for each of the beta coefficients in each of the models:

Table 5: Summary of P-values From Each Model

	Complete.Cases	IPW	Reg.Imputation
(Intercept)	0.086	0.073	0.445
brate	0.146	0.300	0.124
basewt	0.0005	0.001	0.001
s12wgt	0.0001	0.0004	0.0003

We see that the p-values for each of the covariates are not much different between the models.

Question 2:

This analysis asks you to calculate average weight trajectory for each treatment group. What I mean by weight trajectory is mean weight (and standard error) at baseline, week 6, and week 12.

a.) Calculate the weight trajectory using complete cases only.

Table 6: Average Weight Trajectories Using Regression

	Dependent variable:
	weight
s06wgt	1.738
	(4.323)
s12wgt	3.081
	(4.645)
Z	6.149
	(4.022)
s06wgt * Z	0.692
	(6.182)
s12wgt * Z	0.810
	(6.669)
Constant	144.721***
	(2.780)
Observations	569
R^2	0.013
Adjusted R^2	0.004
Residual Std. Error	31.577 (df = 563)
F Statistic	1.482 (df = 5; 563)
Note:	*p<0.1; **p<0.05; ***p<0.01

b.) Calculate the weight trajectory using IPW, with sequential inverse probability weights. To calculate the weight for the week 6 weight, you can use any of the baseline covariates; for the week 12 weight, you can use any of the baseline covariates, and week 6 weight. Smoking status at week 12 cannot be used to calculate the weights.

Table 7: Average Weight Trajectories Using IPW Regression

	Dependent variable:
	weight
s06wgt	2.687
	(5.413)
s12wgt	3.307
	(5.413)
Z	3.612
	(5.528)
s06wgt * Z	0.395
	(7.818)
s12wgt * Z	1.399
	(7.818)
Constant	146.584***
	(3.827)
Observations	423
R^2	0.007
Adjusted R^2	−0.005
Residual Std. Error	43.364 (df = 417)
F Statistic	0.576 (df = 5; 417)
Note:	*p<0.1; **p<0.05; ***p<0.01

c.) Calculate the weight trajectory using sequential imputation. Use either multiple imputation or regression imputation.