PHP 2517 HW#3

Blain Morin

March 16, 2019

Question 1

1a.) Using the output from Model M1, construct a table with estimates and standard errors for each of the parameters in the model. For the parameter, you do not have to write the standard error.

Table 1: M1 Sumary

	Estimate	SE
Fixed Effects		
(Intercept)	-4.628	1.018
Z	1.883	0.982
totfager.cent	-0.766	0.261
postweek	0.057	0.045
Random Effect		
Tau	5.944	NA

1b.) Using the model output, provide an estimate of smoking cessation at week 5 for an individual having i = 0, F = 0, and Z = 0. Do the same for an individual having i = 0, F = 0, and Z = 1.

Table 2: Probability of Cessation for Exercise vs. Control

postweek	totfager.cent	Z	id	Probability
0	0	0	control	0.0097768
0	0	1	exercise	0.0642803

1c.) The R program contains code to construct two histograms, one of the alphai hat and another of a variable pi hat. What quantities are being depicted in each of these histograms? Please be specific with your answer, referring to relevant populations defined by the covariates as appropriate.

$\widehat{\alpha_i}$ are the predicted log odds for each individual if they were to have an average Fagerstrom score and be in the control group at week 5. It can be interpretted as an individual's baseline relative propensity to quit.

$\widehat{p_i}$ are the predicted probabilities of quitting smoking for each individual if they were to have an average Fagerstrom score and be in the control group at week 5. It can be interpretted as an individual's underlying baseline probability of quitting.

Table 3: Subject Specific vs. Population Average Effects

	Sub_Spec	Pop_Ave
1	-4.6277405	-1.2725597
2	1.8832377	0.5178623
3	-0.7662910	-0.2107186
4	0.0567541	0.0156065

1d.) Histograms like the ones in the previous question can sometimes be helpful in assessing whether the normality assumption makes sense for the random intercept. Based on these histograms, give your assessment of the validity of the normality assumption, and provide a brief justification.

The histogram for $\widehat{\alpha_i}$ does not exhibit the normal distribution bell shape. This is evidence that our normality assumption is violated. We further see from the predicted probability of smoking cessation histogram that many of the predictions are near 0. The probabilities are heavily right skewed, which is further evidence that our normaility assumption does not hold.

1e.) Provide an interpretation of beta3, the coefficient of Z from this model.

$\beta_3 = 1.883$ is the average increase in log odds of quitting for people in the exercise group compared to the control group, all else equal. $exp(1.883) = 6.57$. On average, the odds of quitting smoking for an individual in the exercise group are 6.57 times higher than in the control group, all else equal.

1f.) Provide an interpretation of the coefficient beta0.

$\beta_0 = -4.628$ is the average log odds of quitting for people in the control group at week 5 with an average Fagerstrom score. $logit^{-1}(-4.628) = .01$. The average probability of quitting for an individual in the control group at week 5 with an average Fagerstrom score is .01.

1g.) Using the approximation discussed in class, convert the coefficients of Z, F, and t to their "population-averaged" counterparts.

```
divisor = (1+.346*5.944^2)^.5

one.g = data.frame(Sub_Spec = dM1$coef)
one.g = one.g %>%
  mutate(Pop_Ave = Sub_Spec / divisor)

kable(one.g, 'latex', row.names = TRUE, booktabs = TRUE, caption = "Subject Specific vs. Population Ave
```