PHP 2517 HW#3

Blain Morin

March 20, 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 alpha i = 0, F = 0, and Z = 0. Do the same for an individual having alpha 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	0.0097
0	0	1	exercise	0.0604

1c.) The R program contains code to construct two histograms, one of the alpha i hat and another of a variable p i 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 interpreted 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 interpreted as an individual's underlying baseline probability of quitting.

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 normality assumption does not hold.

1e.) Provide an interpretation of beta 3, the coefficient of Z from this model.

$\beta_3 = 1.883$ is the average increase in log odds of quitting for an individual 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 an individual 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.

I calculated the population averaged effect, β^*, using the following formula:

$$\beta^* = \frac{\beta}{(1 + .346(5.944^2))^{1/2}}$$

Table 3: Subject Specific vs. Population Average Effects

	Sub_Spec	Pop_Ave
Z	1.8832	0.5179
totfager.cent	-0.7663	-0.2107
postweek	0.0568	0.0156

Question 2

2a.) Which of the GLM correlation structures in the same as for the multilevel model?

The exchangeable correlation structure corresponds to the random intercept multilevel model.

2b.) Construct a table where the row entries are the regression parameters and the scale parameter phi and the columns correspond to each model. Fill the table with the estimate and standard error of each parameter.

Table 4: Comparing Correlation Structures

	Independant		Exchangeable		Unstructured	
	Estimate	Std.err	Estimate	Std.err	Estimate	Std.err
Coefficients						
(Intercept)	-0.999	0.207	-0.949	0.204	-0.954	0.194
Z	0.549	0.261	0.465	0.258	0.361	0.250
postweek	0.035	0.021	0.015	0.020	0.022	0.018
totfager.cent	-0.199	0.067	-0.183	0.067	-0.158	0.064
Scale Parameter						
phi	1.002	0.063	1.011	0.068	1.016	0.066

2c.) What is the interpretation of gamma3 in model G1.unst?

$\gamma_3 = .361$ is the average increase in log odds of quitting for the exercise group compared to the control group, all else equal. $exp(.361) = 1.43$. On average, being in the exercise group is associated with 1.43 times the odds of quitting smoking compared to being in the control group, all else equal.

2d.) Which of these models gives an estimate of gamma3 that is closest to the *marginalized* estimate of beta3 from the multilevel model?

The independent correlation structure is closest to our marginalized estimate from the multilevel model.

2e.) Construct another table where the row entries correspond to estimates of P(Yij = 1) when (a) tj = 0, F = 0 and Z = 0; and (b) tj = 0, F = 0 and Z = 1. The column entries correspond to each model.

Table 5: Predicted Probability of Quitting for Each Cor Structure

postweek	totfager.cent	Z	id	Independent	Exchangeable	Unstructured
0	0	0	control	0.2692	0.2791	0.2780
0	0	1	exercise	0.3895	0.3814	0.3559

2f.) Provide an explanation as to why these estimates are different from each other.

The estimates are different between the models because the correlation structures are different. In effect, each model is making slightly different comparisons. For the model with an independent correlation structure, each observation is assumed to be independent. Therefore, the model compares the exercise group to the control group, but ignores any time trend or within person correlation. The exchangeable model allows for the within person correlation, but does not fully capture the possible time trend (In other words, the probability of smoking cessation is independent of the previous cessation measurement). The unstructured correlation captures both the within person correlation and time trend.

The estimates between the models are different due to the way they "borrow strength." The independent model does not borrow any strength (people with few observation are given the same amount of weight). The exchangeable model borrows some strength because it estimates the correlation between measurements using all available cessation data. The unstructured model borrows the most strength because it estimates using all cessation data as well as time dependent correlation.

This "borrowing of strength" is seen in Table 5. The difference in the estimated probability is highest in the independent correlation structure model, less in the exchangeable model, and lease in the unstructured model. The more the model "borrows strength," the more its estimates will attenuate toward the mean.

2g.) If you were reporting the results of this study for a journal, which model(s) would you report and why, and what would be your conclusion about treatment effect?

I would report the exchangeable and unstructured correlation structure models. It makes intuitive sense that the model should account for within person correlation (therefore the independent model is inappropriate). Choosing between the exchangeable and unstructured models is more difficult. While it is feasable that quitting smoking one week increases your chances of continuing not to smoke the next week, this may be a strong assumption that needs further analysis. However, we see from Table 4 that the treatment effect is most precise (lowest standard error) in the unstructured model.

Although the treatment effect was not statistically significant at the .05 level in the unstructured correlation structure model, there is some weak evidence that exerecise helps people quit smoking when using the exchangeable model. Since exercise is healthy and may help people quit smoking, I would recommend that further research be conducted.