Bios 767 Homework 7

Problem 1

Descriptive statistics for the toenail data are provided in both a graphical and tabular format below. These descriptives highlight that the Itraconazole treatment group has a higher proportion of patients reporting moderate or severe onycholysis at each of the seven visits.

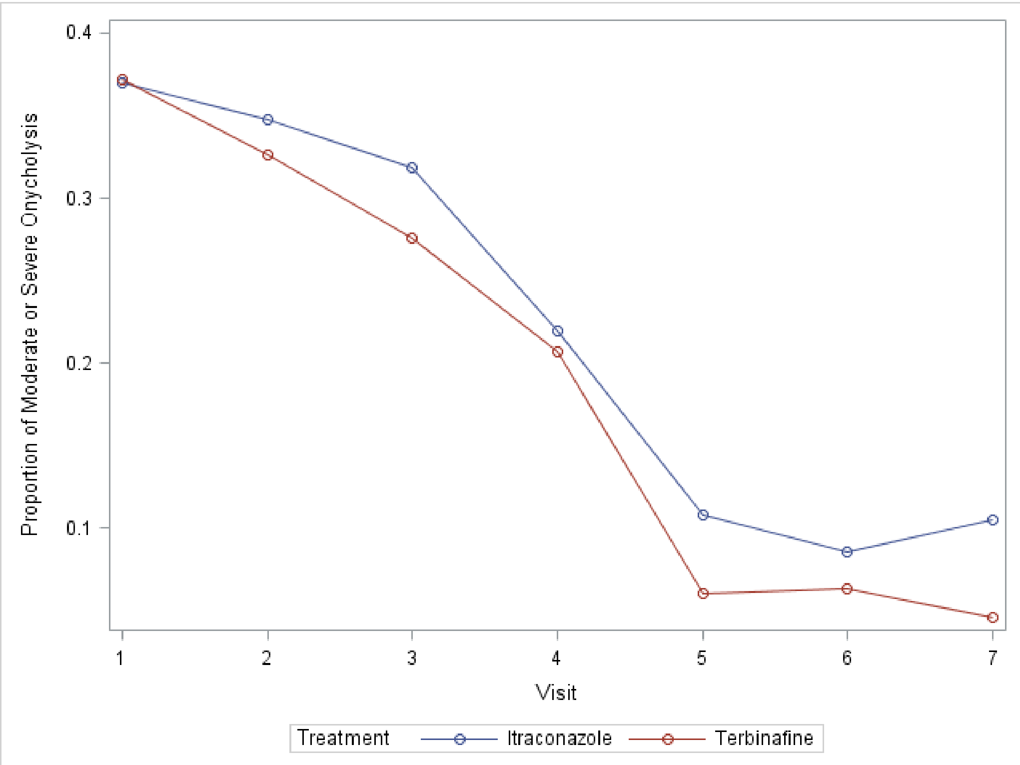

Figure 1: Proportion of patients experiencing moderate or severe onycholysis at visits 1-7 in the two treatment groups.

Table 1: Proportion of patients experiencing moderate or severe onycholysis at visits 1-7 in the two treatment groups.

	Proportion (N)						
Treatment	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7
Itraconazole	0.370 (146)	0.348 (141)	0.319 (138)	0.220 (132)	0.108 (130)	0.085 (117)	0.105 (133)
Terbinafine	0.372 (148)	0.327 (147)	0.276 (145)	0.207 (140)	0.060 (133)	0.063 (127)	0.046 (131)

Problem 2

The sample correlation matrix for the responses across visits 1 - 7 for the Itraconazole treatment group is given by

$$\begin{pmatrix} 1 & 0.8930 & 0.6552 & 0.5012 & 0.4176 & 0.2742 & 0.1830 \\ 0.8930 & 1 & 0.7261 & 0.5707 & 0.4850 & 0.2941 & 0.2787 \\ 0.6552 & 0.7261 & 1 & 0.7244 & 0.4750 & 0.2665 & 0.2401 \\ 0.5012 & 0.5707 & 0.7244 & 1 & 0.5690 & 0.3603 & 0.3235 \\ 0.4176 & 0.4850 & 0.4750 & 0.5690 & 1 & 0.6264 & 0.4026 \\ 0.2742 & 0.2941 & 0.2665 & 0.3603 & 0.6264 & 1 & 0.6717 \\ 0.1830 & 0.2787 & 0.2401 & 0.3235 & 0.4026 & 0.6717 & 1 \end{pmatrix}$$

The sample correlation matrix for the responses across visits 1 - 7 for the Terbinafine treatment group is given by

$$\begin{pmatrix} 1 & 0.8406 & 0.7055 & 0.5190 & 0.2099 & 0.1274 & 0.0570 \\ 0.8406 & 1 & 0.8253 & 0.6295 & 0.2479 & 0.1569 & 0.0801 \\ 0.7055 & 0.8253 & 1 & 0.8042 & 0.3105 & 0.1777 & 0.1014 \\ 0.5190 & 0.6295 & 0.8042 & 1 & 0.4326 & 0.2523 & 0.1650 \\ 0.2099 & 0.2479 & 0.3105 & 0.4326 & 1 & 0.7887 & 0.6496 \\ 0.1274 & 0.1569 & 0.1777 & 0.2523 & 0.7887 & 1 & 0.7591 \\ 0.0570 & 0.0801 & 0.1014 & 0.1650 & 0.6496 & 0.7591 & 1 \end{pmatrix}$$

These sample correlations are computed using an estimate of Pearson's correlation coefficient:

$$r_{Y_{ij}Y_{ik}} = \hat{\rho}_{Y_{ij}Y_{ik}} = \frac{\widehat{Cov}(Y_{ij}, Y_{ik})}{\sqrt{\widehat{Var}(Y_{ij})\widehat{Var}(Y_{ik})}}, \qquad j, k = 1, ..., 7$$

where

$$\widehat{Cov}(Y_{ij}, Y_{ik}) = \frac{1}{N-1}\sum_{i=1}^{N}(Y_{ij} - \bar{Y_j})(Y_{ik} - \bar{Y_k}),$$

$$\widehat{Var}(Y_{ij}) = \frac{1}{N-1}\sum_{i=1}^{N}(Y_{ij} - \bar{Y_j}),$$

and

$$\widehat{Var}(Y_{ik}) = \frac{1}{N-1}\sum_{i=1}^{N}(Y_{ik} - \bar{Y_k}).$$

N is the number of patients with a non-missing response at both visit i and visit j.

For both treatment groups, we see that as the time separation increases, the correlation between the responses at the corresponding visits decreases. We see that the correlations are all positive, which is expected.

Problem 3

The observed log-odds ratios of moderate or severe onycholysis between visit k and visit j, where $k > j$, are presented below for each treatment group separately. Odds ratios were computed using the $2x2$ contingency table corresponding to the two visit of interest, and then the log of the odds ratios were taken. Two log(OR) could not be computed, since one of the cells in the corresponding 2x2 tables had a count of 0.

Table 2: Log-odds ratios between visits for Itraconazole treatment group.

	Y_{i1}	Y_{i2}	Y_{i3}	Y_{i4}	Y_{i5}	Y_{i6}	Y_{i7}
Y_{i1}	0	6.014	3.326	2.706	3.487	2.108	1.194
Y_{i2}		0	3.855	3.183	-	2.225	1.961
Y_{i3}			0	4.526	3.784	1.872	1.534
Y_{i4}				0	4.275	2.411	2.039
Y_{i5}					0	4.605	2.603
Y_{i6}						0	4.383
Y_{i7}							0

Table 3: Log-odds ratios between visits for Terbinafine treatment group.

	Y_{i1}	Y_{i2}	Y_{i3}	Y_{i4}	Y_{i5}	Y_{i6}	Y_{i7}
Y_{i1}	0	5.438	4.156	2.968	1.816	1.044	0.541
Y_{i2}		0	5.312	3.737	2.083	1.268	0.754
Y_{i3}			0	-	2.944	1.415	0.945
Y_{i4}				0	3.646	1.929	1.480
Y_{i5}					0	5.808	4.745
Y_{i6}						0	5.670
Y_{i7}							0

In both treatment groups, the log odds ratio of moderate to severe onycholysis between visit k and visit j, where $k > j$, decreases as the time separation increases. This makes sense since our correlations also decrease as the time separation increases.

Additionally, all of the log-odds ratios are larger than 0; equivalently, all of the odds ratios are larger than 1. This makes sense, since

$$OR(Y_{ik}, Y_{ij}) = \frac{odds(Pr(Y_{ik} = 1|Y_{ij} = 1))}{odds(Pr(Y_{ik} = 1|Y_{ij} = 0))}.$$

The odds of a patient having moderate or severe onycholysis at visit k given that they had moderate or severe onycholysis at a previous visit j will always be larger than the odds of a patient having moderate or severe onycholysis at visit k given that they had no or mild onycholysis at a previous visit j.

Problem 4

Due to the fact that the correlations decline over time as the separation between pairs of repeated measures increases, it seems reasonable to specify an autoregressive structure to the correlation matrix. Do note that after visit 4, the timing of response measurements is spaced three months apart, rather than one month apart.

Problem 5

Due to the fact that the log odds ratios decline over time as the separation between pairs of repeated measures increases, it seems reasonable to specify an autoregressive structure to the log odds ratios.

Problem 6

If we assume an autoregressive structure on the within-subject correlation matrix, then the first diagonal elements off the 1's on the main diagonal are of the form $\rho^{|t_j - t_k|}$, where t_j denotes a visit at the j^{th} month and t_k denotes a visit at the k^{th} month. We will take a weighted average of these 1-step correlations across both treatment groups to obtain an estimate of $\hat{\rho}$. It will be easier to obtain an estimate for $\log(\hat{\rho})$ first.

The sum of $|t_j - t_k| \log(r_{Y_{ij},Y_{ik}})$ for the Itraconazole treatment group is $1\log(0.8930) + 1\log(0.7261) + 1\log(0.7244) + 3\log(0.5690) + 3\log(0.6264) + 3\log(0.6717) = -5.0444$. Similarly, the sum of $|t_j - t_k| \log(r_{Y_{ij},Y_{ik}})$ for the Terbinafine treatment group is -4.6364.

The weighted average is therefore $\frac{-5.0444-4.6364}{(1+1+1+3+3+3)+(1+1+1+3+3+3)} = -0.4034$, so take our estimate as $\log(\hat{\rho}) = -0.4034$. Exponentiating this, we obtain an estimate $\hat{\rho} = 0.6681$.

Under $H_0 : \beta_3 = 0$, $\sqrt{n}(\hat{\beta}_3 - 0) \longrightarrow_d N(0, \lambda_0^2)$.
Under $H_1 : \beta_3 = -\log(1.5)$, $\sqrt{n}(\hat{\beta}_3 - (-\log(1.5))) \longrightarrow_d N(0, \lambda_1^2)$.

We find consistent estimators of λ_0 and λ_1 to be $\hat{\lambda}_0 = 1.0153$ and $\hat{\lambda}_1 = 1.2533$.

Specifying the type I error to be $\alpha = 0.05$ and power to be $1 - \gamma = 0.8$ (so the type II error is $\gamma = 0.2$), we can use the following formula to solve for the sample size in each treatment group:

$$\begin{aligned}
\sqrt{n} &= (Z_{1-\alpha/2}\lambda_0/\lambda_1 + Z_{1-\gamma})/\{(-\log(1.5) - 0)/\lambda_1\} \\
&= (1.96(1.0153/1.2533) + 0.8416212)/\{-\log(1.5)/1.2533\} \\
&= -7.509 \\
\implies n &= 56.39 \approx 57 \\
\implies 2n &\approx 114
\end{aligned}$$

We require a sample size of at 114, or 57 patients per treatment group, in order to achieve 80% power to detect a difference of $-\log(1.5)$ when we set a significance level of $\alpha = 0.05$.