STAT 357 PROJECT

Regression Analysis - Parkinson’s Disease

PRAHAR IJNER

1. Introduction

Parkinson's Disease (PD) is a neurodegenerative disorder characterized by loss or disruption of motor functions with symptoms of tremors, impaired balance, and rigidity of the muscles. In certain cases, people with PD may also experience fatigue, soft speech, and sleep disturbances. Patients with symptoms are clinically diagnosed and given a score on the Unified Parkinson's Disease Rating Scale (UPDRS). The UPDRS score is clinically based on four parts. Part 1 – Mentation, Behavior and Mood; Part 2 – Activities of Daily Living; Part 3 – Motor; Part 4 – Complications. The goal of this paper is to assess various biomedical voice measurements collected from people with early stage Parkinson's disease, and develop a model to predict the clinician's motor UPDRS score. The dataset used for this – parkinsons_motor_updrs.dat[1] consists of 16 biomedical voice measures, a unique identifier for each subject, age, sex, time since the subject was recruited into trial, and the motor UPDRS score by the clinician.

The dataset has several jitter and shimmer measurements, which measure frequency and amplitude instability respectively. Larger jitter and shimmer values show greater instabilities in speech, indicating a lower level of motor control and consequently a higher motor UPDRS score. PPE is another measure of variation in frequency, and we would expect it to have a increasing effect on motor UPDRS, similar to jitter. NHR and HNR are inversely proportional values reflecting the presence of noise. A low NHR and high HNR signals a superior voice quality [1]. We would expect our response to have a positive relationship NHR and negative relationship with HNR. RPDE and DFA previously been used to detect abnormalities in voice measurements [4], however their non-linear nature makes it difficult to predict their effect on the response variable.

[1] See table A1 in appendix A for description of variables

1.1 Gender differences

Epidemiological studies have highlighted several indications of differences of prevalence and incidence of PD. A male subject is 1.5 times more likely to have PD compared to a female subject [_3]. Furthermore, the verbal fluency is impaired to a greater extent in male subjects than female subjects, who suffer greater deficits in visuospatial cognition [5]. In context of our dataset, we would expect lower jitter and shimmer values from female subjects compared to male subjects, indicating lower motor UPDRS scores for women.

1.2 Age variation

Age is considered as one of the main risk factors for PD, with most cases observed in people over the age of 60 years. The severity of motor features of PD increases with and increase in age of onset [2]. Our dataset looks at ages 36 to 85 years. Thus, we would expect the older age group to have a lower verbal fluency due to a combination of PD and the natural aging process. This would result in a positive relation between age and motor UPDRS rating.

1.3 Effect of test time

Parkinson's disease is progressive condition and symptoms get worse with increase in time [6]. Thus, the voice measurements would indicate a better verbal fluency if they were taken 10 days after diagnosis when compared to 100 days after diagnosis. This would imply a negative relation between test time and the motor UPDRS rating. However, the progression of PD varies from person to person. Studies show that increase in age increases the rate of motor progression in patients with PD [7]. We would expect an increased effect of test time on the prediction of motor UPDRS score with an increase in age.

2. Data

Our data has 21 feature measurements collected from 210 subjects. The variable subject is a discrete unique numeric identifier for each subject. The age is the subject's age in number of years. As we aren't considering months, days, etc., this is also a discrete variable. Our final discrete variable is the sex of the participant, which is a binary variable (0 for male, 1 for female).

Feature	min	max	mean	SD
subject	4.00E+00	6.00E+03	3.22E+03	1.83E+03
age	3.60E+01	8.50E+01	6.44E+01	9.15E+00
sex	0.00E+00	1.00E+00	3.33E-01	4.73E-01
test_time	4.99E-01	2.09E+02	9.00E+01	5.64E+01
motor_UPDRS	5.14E+00	3.95E+01	2.13E+01	8.36E+00
Jitter.Percent	1.67E-03	3.59E-02	6.18E-03	4.72E-03
Jitter.Abs	6.51E-06	3.89E-04	4.45E-05	3.95E-05
Jitter.RAP	6.90E-04	1.98E-02	3.05E-03	2.61E-03
Jitter.PPQ5	8.80E-04	2.98E-02	3.24E-03	2.94E-03
Jitter.DDP	2.08E-03	5.95E-02	9.14E-03	7.82E-03
Shimmer	8.30E-03	1.76E-01	3.35E-02	2.55E-02
Shimmer.dB.	8.20E-02	1.51E+00	3.06E-01	2.25E-01
Shimmer.APQ3	3.60E-03	1.01E-01	1.70E-02	1.37E-02
Shimmer.APQ5	4.17E-03	1.16E-01	1.97E-02	1.62E-02
Shimmer.APQ11	6.47E-03	1.30E-01	2.67E-02	1.94E-02
Shimmer.DDA	1.08E-02	3.02E-01	5.11E-02	4.11E-02
NHR	1.58E-03	3.93E-01	3.12E-02	4.21E-02
HNR	4.51E+00	3.25E+01	2.19E+01	4.35E+00
RPDE	2.77E-01	7.84E-01	5.41E-01	9.97E-02
DFA	5.20E-01	8.14E-01	6.47E-01	7.38E-02
PPE	6.31E-02	5.34E-01	2.21E-01	9.26E-02

Table 1 Summary of measurements collected (n = 210)

The jitter measurements are highly correlated with each other as they are all measures for the variation in fundamental frequency. The same can be said about the shimmer measurements which measure the variation in amplitude. This is confirmed by calculating their correlation matrices[2].

[2] See Table B1.a and TableB1.b in appendix B

After performing preliminary analysis on our dataset to observe the effect of each of each individual measurement on the motor UPDRS score, we have the following coefficients (table 2)[3]. The variables subject has been excluded from the dataset as it is a unique identifier which doesn't provide any information about the patient.

Feature	β_1 (coefficient)	β_0 (intercept)	R squared
age	0.23 ***	6.70 .	0.06
sex	-0.22	21.40 ***	0.00
test_time	0.03 **	18.50 ***	0.04
Jitter.Percent	274.69 *	19.63 ***	0.02
Jitter.Abs	20478.79	20.41 ***	0.01
Jitter.RAP	480.20 *	19.86 ***	0.02
Jitter.PPQ5	373.93 .	20.11 ***	0.02
Jitter.DDP	160.03 *	19.86 ***	0.02
Shimmer	46.56 *	19.76 ***	0.02
Shimmer.dB.	5.55 *	19.63 ***	0.02
Shimmer.APQ3	78.65 .	19.98 ***	0.02
Shimmer.APQ5	73.66 *	19.87 ***	0.02
Shimmer.APQ11	79.30 **	19.21 ***	0.03
Shimmer.DDA	26.22 .	19.98 ***	0.02
NHR	25.30 .	20.54 ***	0.02
HNR	-0.46 ***	31.36 ***	0.06
RPDE	13.43 *	14.06 ***	0.03
DFA	-17.76 *	32.80 ***	0.02
PPE	18.46 **	17.24 ***	0.04

Table 2 Model summaries for individual predictor linear models

The age of the subject has a small, yet significant increasing affect on the motor UPDRS score, which was what we expected. Sex, being a binary categorical variable, doesn't affect the motor UPDRS score by itself. The test time was expected to have a negative (decreasing) effect on motor UPDRS but has a small positive effect instead. The jitter, shimmer, NHR, and PPE measurements have a positive effect on the motor UPDRS score, as expected. Despite the jitter measurements

[3] Significance codes: (α = 0) '***' , (α = 0.001) '**', (α = 0.01) '*', (α = 0.05) '.', (α = 0.1) ' '
Same codes are used throughout the paper

being highly correlated with each other, the coefficient of Jitter.Abs isn't significant, while the coefficients of the others are. An increase in RPDE reflects an increase in the response, while an increase in DFA reflects a decrease. Looking at the R squared values for each model, it is clear that none of the measurements can be individually explain any significant amount of variance in the response.

3. Model Selection

For the full model, we drop Jitter.Abs due to its high p value (>> 0.1) for the β_1 coefficient in the preliminary analysis and the fact the it is highly correlated with the other jitter measurements. We retain sex as a predictor despite of its high p value since it may influence verbal fluency in a multi-variable model and is retained for the full model (model A1) as described in section 1.1. The resultant model has an adjusted R^2 of 0.2229, which although better than our individual models, isn't very good. It is interesting to note that NHR and RPDE now have negative effects on the response. Previously, we established the presence of a high degree of correlation between our jitter measurements and between our shimmer measurements. Including all these measurements puts us at the risk of singularities, inaccurate regression coefficients, and inflated standard errors. This could be the reason behind the insignificant p values for the jitter and shimmer measurements. The VIFs (Variance Inflation Factor)[4] were calculated for each coefficient in model A1 and we can see that there are some serious problems in model A1; the jitter and shimmer measurements have severely high VIF values implying they can be explained by the other existing variables. We sequentially drop the measurements with highest VIFs, recalculating the VIF in every step, and end up with model B1. The coefficients of Jitter.RAP, Shimmer.APQ3 and Shimmer.DDA are

[4] See Table B2 in appendix B for summary of VIF tests

greatly reduced, with Shimmer.DDA now having a negative coefficient. The other coefficients remain close to their previous values. The variables sex, RPDE, and PPE have coefficients that have large p values (>> 0.1). Before eliminating them from our model, we perform AIC tests to measure the effect of dropping each of them.

	Model A1		Model B1		Model C	
	Coefficient		Coefficient		Coefficient	
Intercept	54.890	***	53.980	***	55.320	***
age	0.147	*	0.123	*	0.135	*
sex	0.552		0.474			
test_time	0.027	**	0.027	**	0.027	**
Jitter.Percent	507.100					
Jitter.RAP	263300.000		1091.656	*	1337.000	***
Jitter.PPQ5	-998.600					
Jitter.DDP	-87430.000					
Shimmer	157.300					
Shimmer.dB.	-21.170					
Shimmer.APQ3	-258800.000		-232.057	*	-259.900	**
Shimmer.APQ5	182.700					
Shimmer.APQ11	188.800		201.481	*	217.700	**
Shimmer.DDA	86150.000					
NHR	-90.510	*	-105.872	***	-103.700	***
HNR	-0.692	**	-0.648	**	-0.723	**
RPDE	-3.038		-0.166			
DFA	-51.280	***	-51.297	***	-48.770	***
PPE	14.720		12.895			
RSE	7.37		7.331		7.3	
df	191		198		201	
Adjusted R2	0.2229		0.2312		0.2377	
AIC	1454.97		1446.281		1441.6	

Table 3 Linear regression variable coefficients and model summary

The AIC[5] values confirm that the variables sex, RPDE, and PPE should be dropped, which leads us to model C. We have now satisfied parsimony without compromising the effectiveness of our

[5] See table B3 in appendix B for AIC test results for model B1

model. However, we still have an underwhelming value for adjusted R^2 and our model diagnostics point out several key issues. Firstly, the normal Q-Q plot indicates a clear deviation from the normality assumption. The residuals vs fitted values plot doesn't show major issues with 66% residuals within 1SD, and 97% within 2SD. The distribution is in a band like pattern centered at 0. Observations 10, 22, and 92 stand out with high values of Cook's D and have a relatively low leverage, due to which we can consider them as outliers.

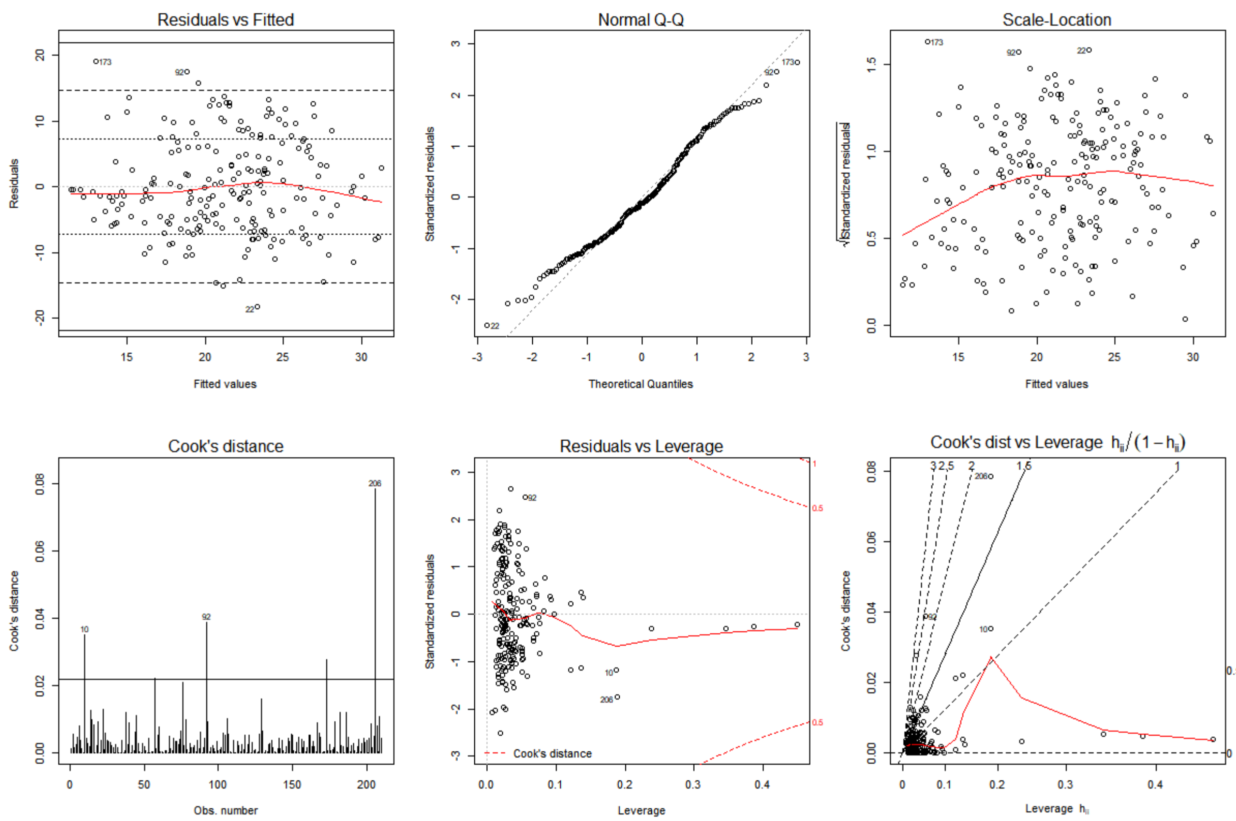

Figure 1 Diagnostic plot for model C

The deviation from normality suggests that we need to transform our data. We use the Box-Cox transformation to determine the optimal transformation for our response variable.

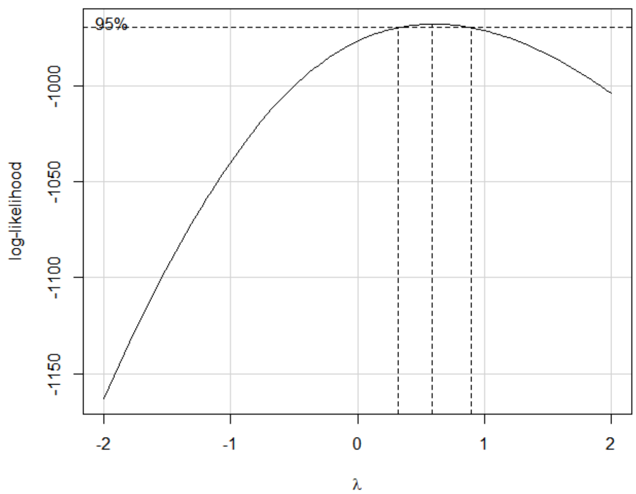

Figure 2 Optimal lambda value for response

This suggests that motor_UPDRS should have a lambda value of approximately 0.5. Since our response doesn't include negative values, we can apply this transformation and calculate a new full model with the applied transformation. We are aware of the high degree of correlation within the jitter values and shimmer values, due to which we have incorrect estimates of our beta coefficients in model D1. It is important to note that we cannot compare the RSE, adjusted R2, and AIC values from our previous (non transformed) models to the new (transformed) model as we have changed the scale. Since we have not transformed any of our predictor variables, their VIF still remains the same, and we can eliminate Jitter.Percent, Jitter.PPQ5, Jitter.DDP, Shimmer, Shimmer.dB., Shimmer.APQ5, and Jitter.PPQ5 as we did in model B. We now have a much better-looking model E1, with the RSE value and AIC decreased, and adjusted R^2 increased. The coefficients of sex, RPDE, and PPE are not significant and to make sure the removal of these variables does not affect our model, we use the perform the AIC[6] drop test and drop the variables

[6] See table B4 in appendix B for AIC test results for model E1

with the highest AIC in every step, until there is a negative effect of removal of a variable to our model.

	Model D		Model E1		Model F		Model G	
	Coefficient		Coefficient		Coefficient		Coefficient	
Intercept	8.452	***	8.373	***	8.627	***	8.915	***
age	0.016	*	0.013	*	0.015	*	0.014	*
sex	0.057		0.046					
test_time	0.003	**	0.003	**	0.003	**	0.003	**
Jitter.Percent	74.310							
Jitter.RAP	29110.000		125.686	*	152.790	***	187.748	***
Jitter.PPQ5	-104.400							
Jitter.DDP	-9678.000							
Shimmer	11.260							
Shimmer.dB.	-2.185							
Shimmer.APQ3	-31000.000		-27.425	*	-30.810	**	-32.614	**
Shimmer.APQ5	14.090							
Shimmer.APQ11	24.680		23.202	**	25.157	**	25.209	**
Shimmer.DDA	10320.000							
NHR	-9.901	*	-11.854	***	-11.617	***	-12.523	***
HNR	-0.080	**	-0.075	**	-0.086	***	-0.086	***
RPDE	-0.155		0.141					
DFA	-5.954	***	-5.989	***	-5.688	***	-6.141	***
PPE	1.414		1.421					
RSE	0.821		0.8169		0.8136		0.8014	
df	191		198		201		198	
Adjusted R2	0.234		0.2424		0.2485		0.277	
AIC	533.428		524.67		520.1188		506.567	

Table 4 Linear regression variable coefficients and model summary after response transformation

The final model (model F) looks similar to the final model prior to transformation in that it uses the same variables to explain the response variable. The final RSE is the lowest of the transformed models at 0.8136, with adjusted R^2 at 0.2485, and AIC of 520.1188. We perform the model diagnostics on model F and observe the distribution of residuals is a lot closer to the normal line. There still is a small deviation from normality at the upper tail, but for the most part it appears to be normal. We have observations 10, 92, and 206 with larger cook's D values, but they have a small leverage due to which we can consider them as outliers. Observations 10 and 92 were also detected as outliers in our previous (pre-transformation) model.

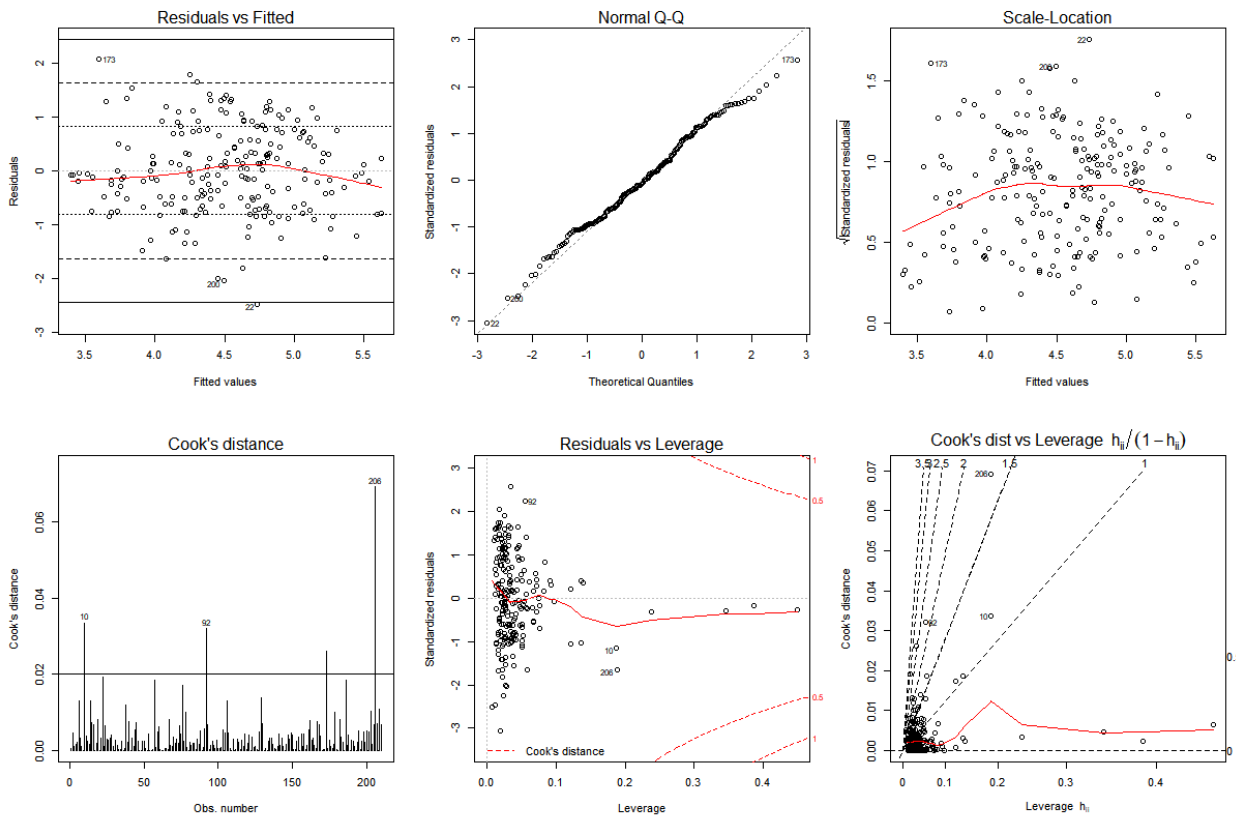

Figure 3 Diagnostic plot for model F

On removal of the outliers from our dataset, we recalculate the model and perform the model diagnostics on it. The revised model (model G) has a much lower AIC and R2, which is due to the fact that we now have a smaller amount of variance in the data after removing the three outliers. The diagnostics still suggest a small deviation from normality in the normal Q-Q plot, but it isn't significant enough to violate the normality assumption. The loess lines of the leverage plots look mostly straight. The homoscedasticity assumption is not violated either. Thus, we select model G as our final model.

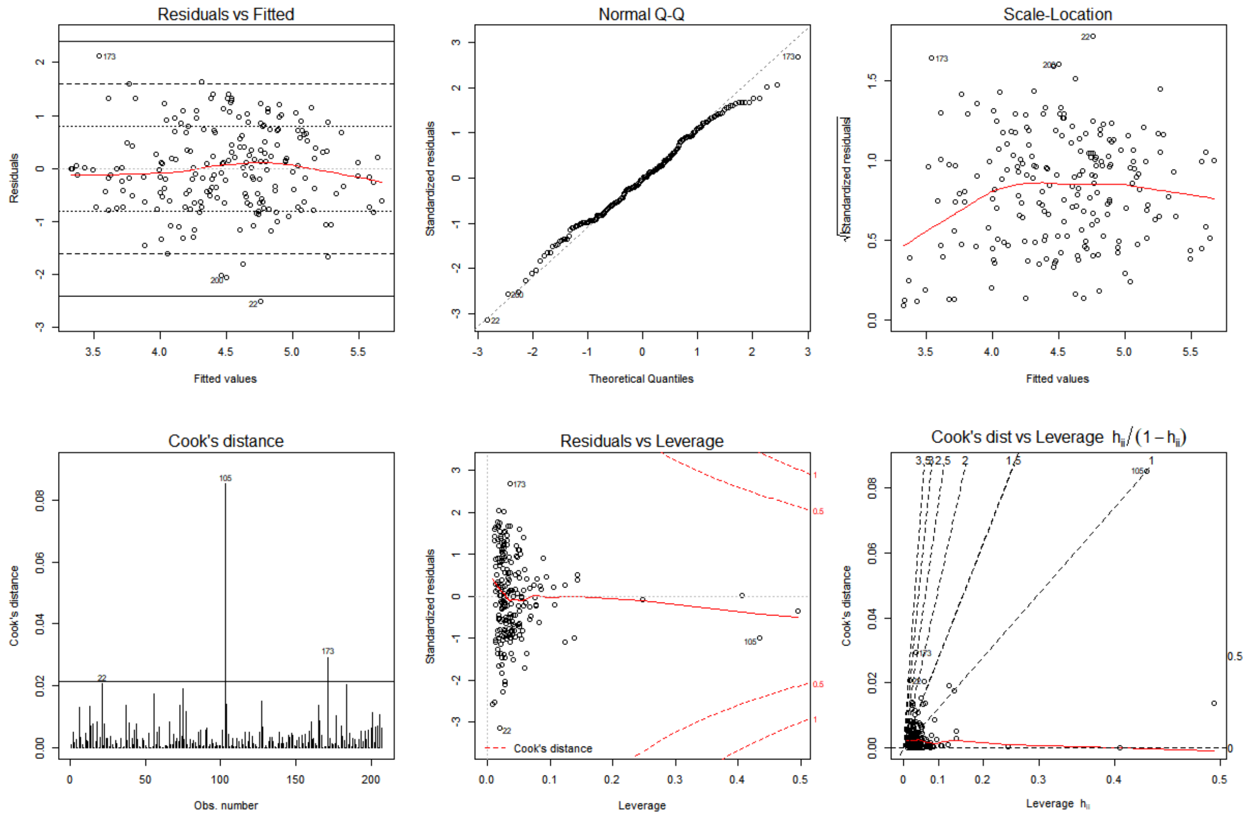

Figure 4 Diagnostic plot for model G

4. Final model

After analyzing all the models, we have selected model G as our final model, which is written in the form:

$$motor\_UPDRS = (8.915 + 0.014 * age + 0.003 * test\_time + 187.748 * Jitter.RAP - 32.614 * Shimmer.APQ3 + 25.209 * Shimmer.APQ11 - 12.523 * NHR - 0.086 * HNR - 6.141 * DFA)^2$$

As we predicted, an increase in age reduces the verbal fluency of a person with PD, and hence increases the motor UPDRS score. Contrary to the assumption of decrease in motor_UPDRS with increase in test time, we observe a very small decrease in the response. This could either be due to misinterpretation of the term test time, meaning it would mean the time after measurements were taken when the motor UPDRS score was given. It could also be due to how different the growth of PD's symptoms are in different people. We dropped all but one jitter measurements due to the high degree of correlation between them. The Jitter.RAP measurement that was retained, increases the motor_UPDRS score as it is increased. Since the Jitter.RAP measurement itself is a small value, it has a large coefficient. This doesn't necessarily mean that Jitter.RAP has the largest influence on the motor UPDRS score. The variables Shimmer.APQ3, NHR, and HNR have decreasing effects on the UPDRS score. Meanwhile Shimmer.APQ11 and DFA have increasing effects on the response. The APQ in shimmer measurements is calculated by the average absolute difference between the amplitude of a period and the average of the amplitudes of its neighbours, divided by the average amplitude. APQ3 is calculates for the direct neighbours, while APQ11 is calculated for the 10 nearest neighbours [8]. This means that a greater difference in the amplitudes over a longer period of time increases the motor UPDRS score, while a difference in amplitudes in a shorter term decreases the score. It is interesting to see that the response in negatively affected by the NHR and HNR measurements as they were expected to increase the score. The effect of DFA was unknown prior to the analysis, and we can see that it has a strong negative effect on the score.

5. Conclusion

The model we have selected does a fair job at predicting the motor UPDRS score by the clinician based on certain voice measurements collected. We say in the previous section how various variables increase and decrease the predicted score. We could have improved the fit of our model by transforming the measurements collected to a different scale, but this may not be appropriate as we don't see severe violations of our model assumptions. Furthermore, using the boxTidwell function from the car library in R also produces large p values (>> 0.1) for transformations for all covariates, implying that they don't need to be transformed.

	MLE of lambda	Score Statistic (z)	Pr(>|z|)
age	-2.82015	-0.4551	0.6491
test_time	1.67343	0.674	0.5003
Jitter.RAP	0.49478	-1.4717	0.1411
Shimmer.APQ3	0.52804	0.5111	0.6093
Shimmer.APQ11	0.90669	0.3504	0.7261
NHR	0.9181	0.9127	0.3614
HNR	5.31807	-1.4924	0.1356
DFA	0.73212	-0.0892	0.9289

Table 5 Result of box Tidwell transformation

It should be noted that the motor UPDRS score is not calculated solely based on verbal fluency of patients. There may be cases where a patient experiences other motor symptom like tremors and rigidity without major vocal symptoms. In this case our model would predict a low motor UPDRS score, when the true score would be a lot higher.

Appendix A

Variable name	Description
subject	Unique integer value that identifies each subject
age	Subject age
sex	Subject gender: '0' - male, '1' - female
test_time	Time since recruitment into trial. The integer part is number of days
motor_UPDRS	Clinician's motor UPDRS score (part 3 of UPDRS test), linearly interpolated
Jitter.Percent	
Jitter.Abs	
Jitter.RAP	(Jitter) Measurements of variation in fundamental frequency
Jitter.PPQ5	
Jitter.DDP	
Shimmer	
Shimmer.dB.	
Shimmer.APQ3	(Shimmer) Measurements of variation in amplitude
Shimmer.APQ5	
Shimmer.APQ11	
Shimmer.DDA	
NHR	Noise to Harmonic Ratio
HNR	Harmonic to Noise Ratio
RPDE	Recurrence Period Density Entropy - nonlinear dynamical complexity measure
DFA	Detrended Fluctuation Analysis - Signal fractal scaling exponent
PPE	Pitch Period Entropy - a non linear measure of fundamental frequency variation

Table A1 Description of all variables in dataset

Appendix B

	Jitter .Percent	Jitter .Abs	Jitter .RAP	Jitter .PPQ5	Jitter .DDP
Jitter.Percent	1.00	0.89	0.98	0.93	0.98
Jitter.Abs	0.89	1.00	0.90	0.72	0.90
Jitter.RAP	0.98	0.90	1.00	0.88	1.00
Jitter.PPQ5	0.93	0.72	0.88	1.00	0.88
Jitter.DDP	0.98	0.90	1.00	0.88	1.00

Table B1.a Correlation between Jitter measurements (all p values << 0.05)

	Shimmer	Shimmer .dB.	Shimmer .APQ3	Shimmer .APQ5	Shimmer .APQ11	Shimmer .DDA
Shimmer	1.00	0.99	0.99	0.98	0.95	0.99
Shimmer.dB.	0.99	1.00	0.97	0.98	0.95	0.97
Shimmer.APQ3	0.99	0.97	1.00	0.96	0.91	1.00
Shimmer.APQ5	0.98	0.98	0.96	1.00	0.96	0.96
Shimmer.APQ11	0.95	0.95	0.91	0.96	1.00	0.91
Shimmer.DDA	0.99	0.97	1.00	0.96	0.91	1.00

Table B1.b Correlation between shimmer measurements (all p values << 0.05)

	VIF							
	Model A1	Model A2	Model A3	Model A4	Model A5	Model A6	Model A7	Model B
age	1.19	1.18	1.16	1.16	1.16	1.15	1.14	1.14
sex	1.23	1.23	1.22	1.22	1.22	1.20	1.20	1.20
test_time	1.07	1.06	1.06	1.06	1.05	1.05	1.05	1.05
Jitter.Percent	71.96	71.96	69.91	69.56	69.56			
Jitter.RAP	1055560.00	1053305.00	44.00	43.58	41.32	7.63	7.41	5.08
Jitter.PPQ5	21.59	21.36	21.29	21.20	16.24	11.82	10.97	
Jitter.DDP	1058026.00	1055810.00						
Shimmer	264.00	264.00	262.94					
Shimmer.dB.	86.59	86.15	85.89	57.53	56.74	54.55		
Shimmer.APQ3	28048690.00	109.97	109.54	49.36	27.91	25.40	8.04	7.59
Shimmer.APQ5	80.99	80.98	80.94	80.33				
Shimmer.APQ11	30.55	30.06	29.96	25.93	17.13	17.02	9.56	8.94
Shimmer.DDA	28048850.00							
NHR	10.46	10.44	10.43	9.77	9.48	9.48	7.85	4.82
HNR	4.82	4.79	4.77	4.73	4.71	4.45	4.45	4.45
RPDE	2.15	2.15	2.13	2.10	2.10	2.07	2.04	1.96
DFA	2.10	2.10	2.09	2.04	2.03	2.03	2.01	1.89
PPE	5.77	5.77	5.76	5.61	5.28	4.48	4.48	4.47

Table B2 VIF test summary to eliminate multicollinearity in model

	Model B1		Model B2		Model B3		Model C	
Variable being dropped	RSS	AIC	RSS	AIC	RSS	AIC	RSS	AIC
RPDE	10641	846.33						
sex	10650	846.5	10650	844.51				
PPE	10708	847.64	10710	845.68	10711	843.71		
<none>	10641	848.33	10641	846.33	10650	844.51	10711	843.71
age	10875	850.9	10878	848.95	10887	847.13	11009	847.47
Shimmer.APQ3	10919	851.75	10920	849.76	10928	847.91	11087	848.96
Jitter.RAP	10974	852.8	10979	850.89	10996	849.22	11151	850.16
Shimmer.APQ11	10999	853.28	10999	851.28	11004	849.36	11184	850.77
HNR	11014	853.55	11053	852.31	11060	850.44	11286	852.68
test_time	11090	855.01	11093	853.07	11113	851.44	11390	854.61
NHR	11503	862.68	11509	860.79	11515	858.9	11573	857.95
DFA	12223	875.44	12225	873.46	12236	871.67	12243	869.77

Table B3 AIC drop test to eliminate predictors from model B1 based on parsimony

	Model E1		Model E2		Model E3		Model F	
Variable being dropped	RSS	AIC	RSS	AIC	RSS	AIC	RSS	AIC
RPDE	132.16	-75.251						
sex	132.22	-75.153	132.23	-77.139				
PPE	132.95	-74.003	133.05	-75.841	133.05	-77.835		
<none>	132.14	-73.284	132.16	-75.251	132.23	-77.139	133.05	-77.835
age	134.78	-71.122	134.9	-72.942	134.97	-74.837	136.55	-74.384
Shimmer.APQ3	136.03	-69.197	136.09	-71.095	136.15	-73.005	138.28	-71.75
Jitter.RAP	136.55	-68.384	136.56	-70.373	136.72	-72.132	138.34	-71.652
Shimmer.APQ11	136.89	-67.866	136.92	-69.822	136.94	-71.785	138.93	-70.763
HNR	137.15	-67.463	137.16	-69.447	137.33	-71.186	141.13	-67.457
test_time	137.16	-67.447	137.97	-68.208	138.02	-70.136	141.92	-66.285
NHR	142.94	-58.784	142.95	-60.765	143.07	-62.592	143.86	-63.442
DFA	153.71	-43.536	153.71	-45.535	153.8	-47.405	153.88	-49.291

Table B4 AIC drop test to eliminate predictors from model E1 (post transformation) based on parsimony

Appendix C

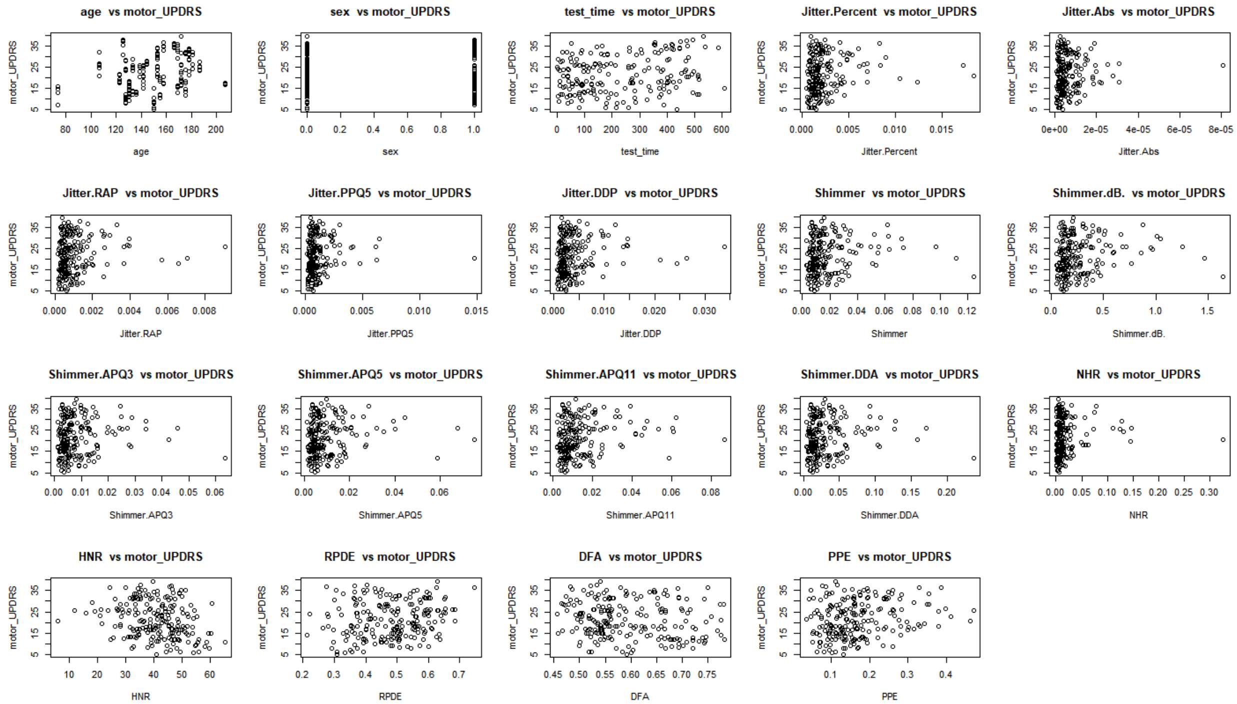

Figure C1 motor_UPDRS vs measurements plots

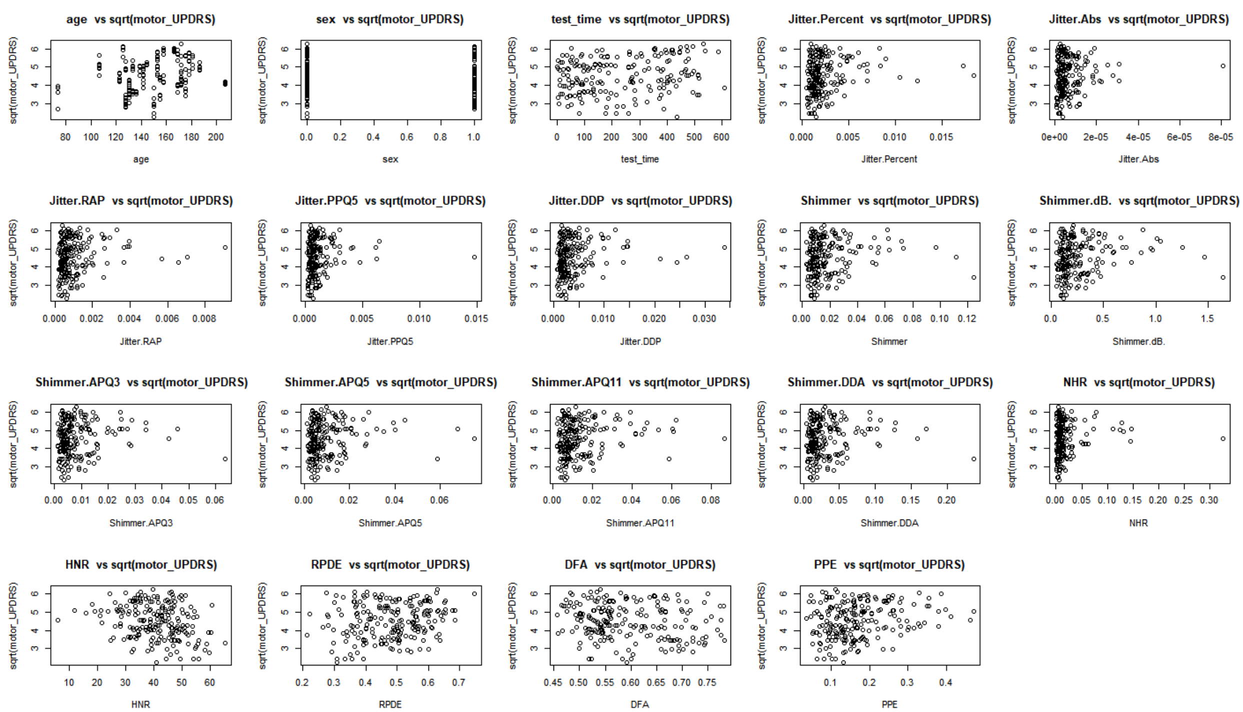

Figure C2 sqrt(motor_UPDRS) vs measurements plots

References

1. Susana Finger L, Aparecida Cielo C, Schwarz K. Acoustic vocal measures in women without voice complaints and with normal larynxes. *Brazilian Journal of Otorhinolaryngology*. 2009;75(3):432-440. doi:10.1016/S1808-8694(15)30663-7

2. Pagano G, Ferrara N, Brooks DJ, Pavese N. Age at onset and Parkinson disease phenotype. *Neurology*. 2016;86(15):1400-1407. doi:10.1212/WNL.0000000000002461

3. Wooten GF, Currie LJ, Bovbjerg VE, Lee JK, Patrie J. Are men at greater risk for Parkinson's disease than women? *J Neurol Neurosurg Psychiatry*. 2004;75(4):637-639. doi:10.1136/jnnp.2003.020982

4. Little MA, McSharry PE, Roberts SJ, Costello DAE, Moroz IM. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *Biomed Eng Online*. 2007;6:23-23. doi:10.1186/1475-925X-6-23

5. Miller IN, Cronin-Golomb A. Gender differences in Parkinson's disease: clinical characteristics and cognition. *Mov Disord*. 2010;25(16):2695-2703. doi:10.1002/mds.23388

6. Progression of Parkinson's. Parkinson Canada. https://www.parkinson.ca/about-parkinsons/progression-of-parkinsons/. Accessed April 17, 2020.

7. Levy G. The Relationship of Parkinson Disease With Aging. *Archives of Neurology*. 2007;64(9):1242-1246. doi:10.1001/archneur.64.9.1242

8. Voice 3. Shimmer. http://www.fon.hum.uva.nl/praat/manual/Voice_3__Shimmer.html. Accessed April 21, 2020.