Name: Adam Hall

Project Title: Can heart rate help predict the average speed achieved during a run?

Section 1 Introduction

As a runner, being able to predict your average speed is useful. There are two obvious ways to do this: distance and time; stride length and cadence. I wanted to explore other explanatory variables to see if they also have a relationship with average speed. This is of particular interest to me because I would like to become a better (faster) runner.

The dataset I am using was collected by me and my fiancé via the use of a Garmin smartwatch. The variables I'll be exploring include "avgspeed", the response variable measured in miles/hour; "avghr", the average heart rate measured in beats/min; "maxhr", the maximum heart rate measured in beats/min; "runner", the person doing the exercise; "course", a qualitative variable determined by elevation change in the course ran.

Section 2 Statistical Analysis

To analyze the data, the statistical program R is used. Significance is observed below .05. Correlation was investigated using a scatterplot matrix. Sample mean and sample variance were collected for all quantitative variables. A linear regression is fitted to the log of the mean vs. log of the variance to determine the variance function for fitting a GLM. Several diagnostic plots are investigated for each possible GLM including: fitted values vs residuals, working residuals vs linear predictor, QQ plot of residuals, and Cook's Distance.

Section 3 Results and Conclusions

Table 1 – Summary statistics for continuous variables by runner.

	Adam		**Leah**	
	Mean	**Variance**	**Mean**	**Variance**
avgspeed	6.59	0.39	5.70	1.79
avghr	157.99	60.54	155.86	329.94
maxhr	179.26	49.02	180.58	98.48
courseDOWNHILL	68		117	
courseFLAT	22		39	
courseUPHILL	1		12	

The average speed of the male runner in this sample is faster and less varied than the female runner. The means for average heart rates are very similar for both runners, but the female runner has significantly more variation that may ultimately affect the model. Max heart rate is very similar to average heart rate with similar means from each but more variation in the female runner. Both runners predominately run courses that are downhill (elevation gain < elevation loss). The female runner tends to log more runs than the male, which may account for some of the variation observed.

Several different break points were investigated before settling on the plot shown in Figure 1 below. For a small number of breaks the slope was greater than 4. Increasing the number of breaks resulted in a slope less than or equal to 0. It should be noted that there is a highly influential point in the bottom left corner of the plot that tended to increase the slope of the line for breaks less than 5. Because of the great variation, investigation of GLMs included Normal, Gamma, Inverse Gaussian and Tweedie. Ultimately, there wasn't one GLM that had significantly better diagnostic plots than the others. The decision was made to use the Tweedie GLM based on the variance function results from the slope observed in Figure 1. The maximum estimated Tweedie index parameter was determined to be 1.5 as shown in Figure 2.

Figure 1 – Plot of the log of group means against the log of group variances.

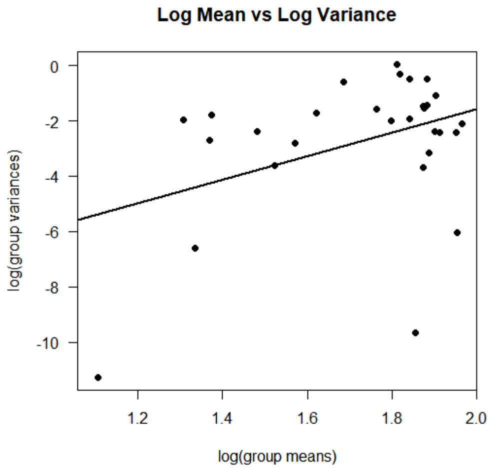
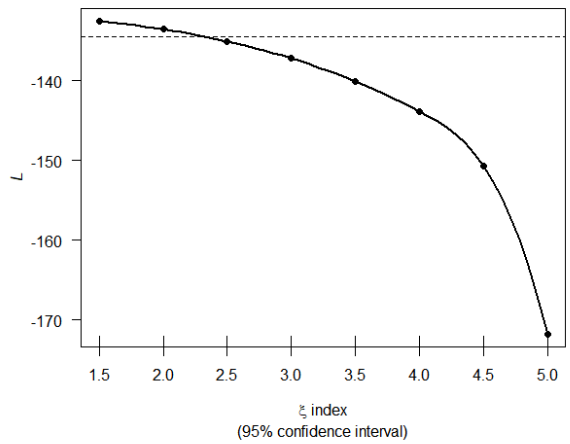
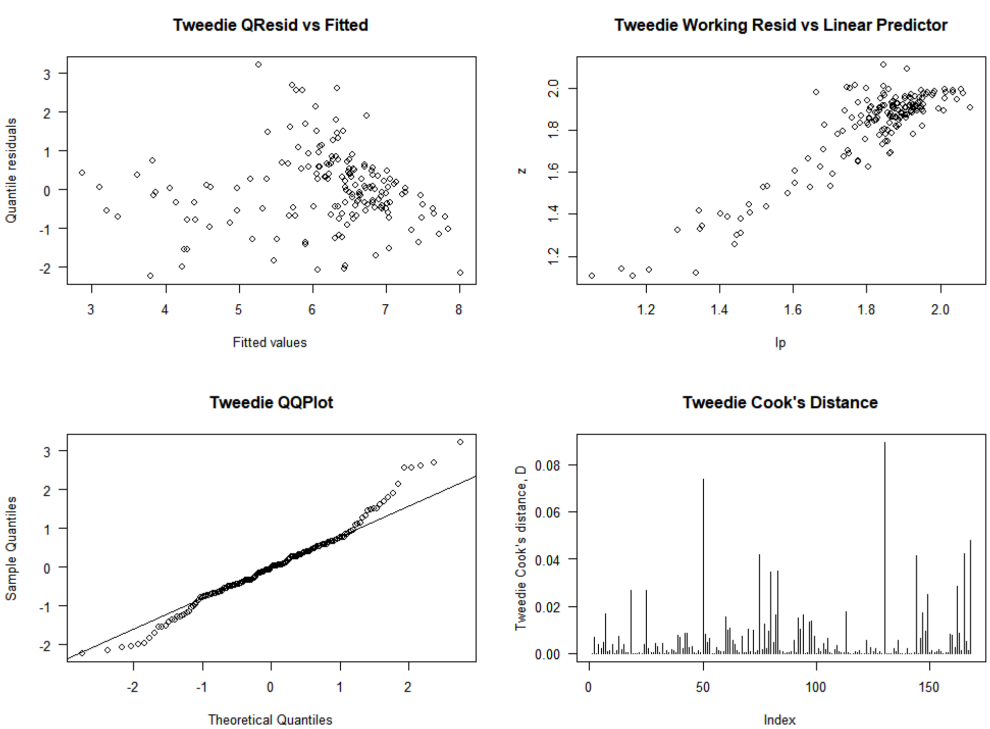


Figure 2 – Maximum likelihood estimate of the Tweedie index parameter power.



As shown below in Figure 3, the diagnostic plots depict a few issues. There are outliers present (based on quantile residuals with an absolute value greater than 2). This is also seen in the QQ Plot of quantile residuals (heavy tailed) which indicates extreme values. There are two points with significantly higher cook's distance compared to the others which may indicate the presence of slightly influential points. The plot of working residuals vs linear predictor looks to be OK.

Figure 3 – Diagnostic plots for Tweedie GLM with log link function.



An investigation into interaction terms concluded that none were significant. The estimated dispersion for the final model was 0.02. Shown below is the final model. Table 2 depicts the summary output and confidence interval.

Average Speed = 0.746 – 0.005*maxhr -0.119*runnerLeah + 0.13*avghr – 0.088*courseFLAT – 0.144*courseUPHILL

Table 2 – Summary Output for final model, including a 95% confidence interval

	Estimate	Standard Error	t-value	Pr(>|t|)	Confidence Interval
Intercept	0.746	0.162	4.593	8.72×10^{-6}	[0.425,1.067]
maxhr	-0.005	0.001	-3.756	0.0002	[-0.007,-0.002]
runnerLeah	-0.119	0.015	-8.133	1.04×10^{-13}	[-0.148,-0.09]
avghr	0.013	0.001	15.357	0	[0.011,-0.053]
courseFLAT	-0.088	0.018	-4.959	1.77×10^{-6}	[-0.122,-0.053]
courseUPHILL	-0.144	0.031	-4.651	6.83×10^{-6}	[-0.205,-0.083]

Section 4 – Discussion

While all of the explanatory variables in the model were found to be significant, it should be noted that all of the coefficients are very close to 0. Normally this could indicate issues with the model, but in this sample it is to be expected as we are trying to fit heart rate (100-200 bpm) to the log of average speed (with all average speeds less than 10 mph). Ultimately my goal was to find a relationship between heart rate and average speed. This model indicates that increasing the average heart rate by 1 bpm will increase the average speed by a factor of 1.013. Physiologically there is an upper limit to how much the human heart can beat in a minute, so I imagine there are diminishing returns here. Pushing too hard seems to have a slightly detrimental effect on average speed as increasing the maximum heart rate by 1 bpm reduces the average speed by a factor of .995.

It would be interesting to reexamine this with new data that includes runs where the average heart rate was increased significantly (sprinting the entire time), more runners were sampled, or where the runner maintained a more relaxed approach (there are about 4-5 data points in this sample where the heart rate stayed very low which creates outliers whereas having more observations could create a bimodal dataset).

As runner conditioning (physical fitness) improves, the average heart rate will lower. Perhaps the model presented here still needs improvement to suggest anything for runners outside the sample.