MATH 204 - ASSIGNMENT 2: SOLUTIONS

We carry out the analysis in three phases:

1. Exploratory analysis using scatterplots to discover which predictor variables are likely to influence the response.
2. Model fitting using the regression and general linear model tools in SPSS to discover which models (which collections of predictors) seem to fit best.
3. Verification of the model fit for the selected model by inspection of residual plots.

For 1: Plots indicate that there is a systematic variation of **Strength** with **Age** that seems intially to be quadratic (Figure 1). However, as suggested in lectures, we need to be on the lookout for lurking subgroups. Here we have two factor predictors **Group** and **AgeGroup** that together define six potential subgroups (in a 3×2 cross-classification), and when a plot of the six subgroups is produced (Figure 2), a different picture emerges: it seems that there is actually a set of straight-line relationships. A plot of **Strength** versus **Test** (Figure 3) reveals no such pattern. The conclusion of this analysis is that a model involving **Group** , **AgeGroup** and **Age** should be sufficient to explain the variation in **Strength** . We might also consider models involving **Age** ^{2}.

For 2: The analyses included in the SPSS are for the several different models. After exploring the model fits of these models, we can conclude that the model

$$\textbf{Group} + \textbf{AgeGroup} + \textbf{Age} + \textbf{AgeGroup} . \textbf{Group}$$

(listed in Analysis 6) provides a good fit ($R^2 = 0.812$, Adjusted $R^2 = 0.804$), and that really there is no need for the quadratic term, although models that include the quadratic term also seem to fit adequately. Parameter estimates for the selected model are presented in the **Parameter Estimates** table.

For 3: A residual plot of the selected model indicates that there is apparently no pattern in the standardized residuals (Figures 4 and 5). The residual plot for the simple quadratic model

$$\textbf{Group} + \textbf{Age} + \textbf{Age}^2$$

is also presented (Figure 6).

In order to get full marks, students should

1. Look at some plots of the raw data

   4 Marks
2. Check the fits of different models, and select one as their chosen model, giving reasons as to why the model was chosen, and comparing R^2 statistics. They need to report the parameter estimates in a table or model formula (as in the textbook) for the model.

   12 Marks
3. Check the adequacy of fit using residual plots.

   4 Marks

Figure 1

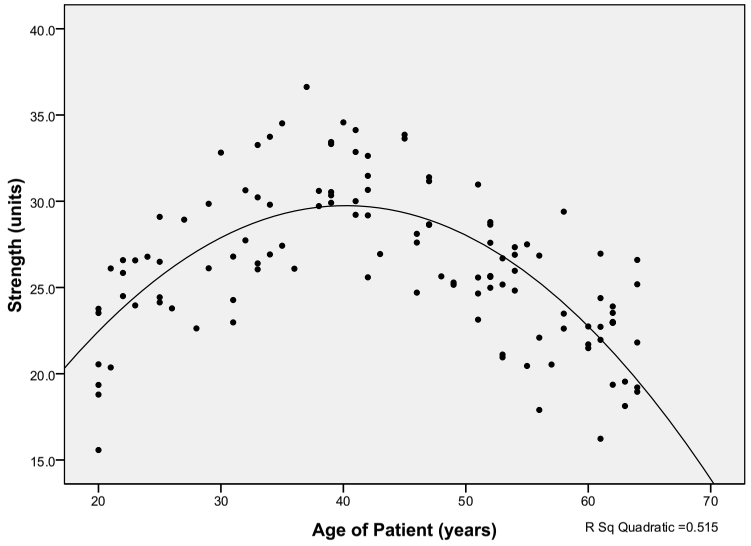


Figure 2

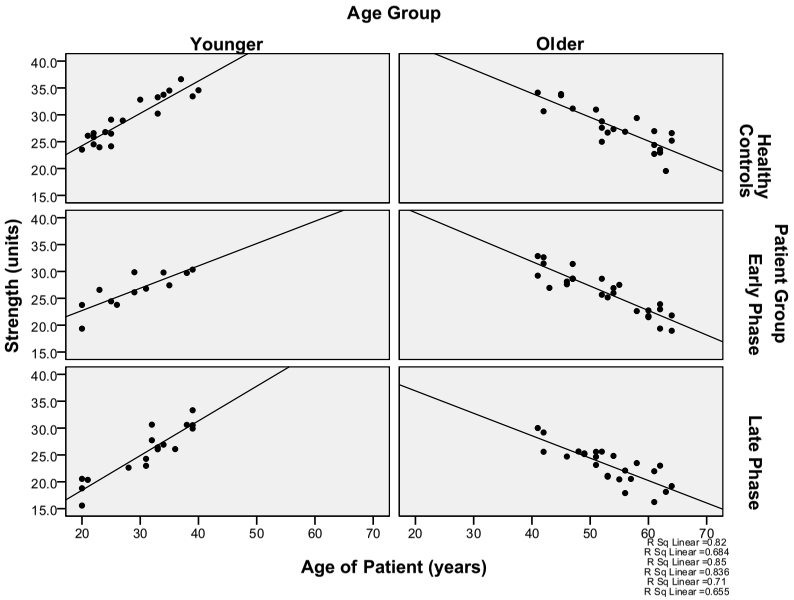
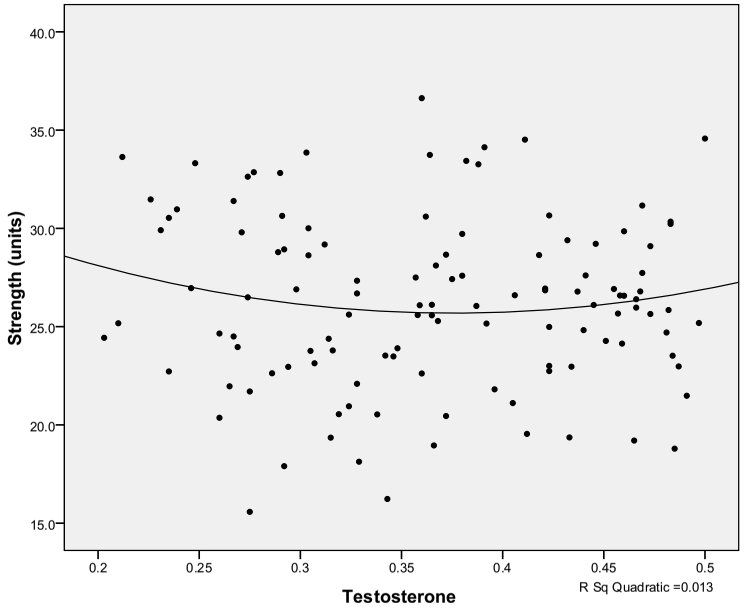


Figure 3



Tests of Between-Subjects Effects

Dependent Variable: Strength (units)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1828.731 [a]	13	140.672	36.591	.000
Intercept	51.218	1	51.218	13.323	.000
Group	13.543	2	6.771	1.761	.177
AgeGroup	32.770	1	32.770	8.524	.004
Age	1.171	1	1.171	.305	.582
AgeSq	1.153	1	1.153	.300	.585
Group * Age	7.276	2	3.638	.946	.392
Group * AgeSq	4.566	2	2.283	.594	.554
AgeGroup * Age	15.072	1	15.072	3.921	.050
AgeGroup * AgeSq	1.888	1	1.888	.491	.485
Group * AgeGroup	15.118	2	7.559	1.966	.145
Error	399.818	104	3.844		
Total	83228.312	118			
Corrected Total	2228.550	117			

a. R Squared = .821 (Adjusted R Squared = .798)

Analysis 1

Tests of Between-Subjects Effects

Dependent Variable: Strength (units)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1822.389 [a]	10	182.239	48.010	.000
Intercept	48.432	1	48.432	12.759	.001
Group	27.937	2	13.969	3.680	.028
AgeGroup	68.573	1	68.573	18.065	.000
Age	.878	1	.878	.231	.631
AgeSq	.326	1	.326	.086	.770
Group * Age	8.677	2	4.338	1.143	.323
AgeGroup * Age	78.259	1	78.259	20.617	.000
Group * AgeGroup	11.767	2	5.883	1.550	.217
Error	406.160	107	3.796		
Total	83228.312	118			
Corrected Total	2228.550	117			

a. R Squared = .818 (Adjusted R Squared = .801)

Analysis 2

Tests of Between-Subjects Effects

Dependent Variable:Strength (units)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1822.064[a]	9	202.452	53.790	.000
Intercept	2764.676	1	2764.676	734.552	.000
Group	28.210	2	14.105	3.748	.027
AgeGroup	1131.353	1	1131.353	300.591	.000
Age	22.707	1	22.707	6.033	.016
Group * Age	8.582	2	4.291	1.140	.324
AgeGroup * Age	1248.035	1	1248.035	331.593	.000
Group * AgeGroup	11.505	2	5.752	1.528	.222
Error	406.486	108	3.764		
Total	83228.312	118			
Corrected Total	2228.550	117			

a. R Squared = .818 (Adjusted R Squared = .802)

Analysis 3

Tests of Between-Subjects Effects

Dependent Variable:Strength (units)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1830.606[a]	11	166.419	44.329	.000
Intercept	2727.728	1	2727.728	726.584	.000
Group	29.583	2	14.791	3.940	.022
AgeGroup	1122.231	1	1122.231	298.928	.000
Age	16.471	1	16.471	4.387	.039
AgeGroup * Age	1213.994	1	1213.994	323.371	.000
Group * Age	14.831	2	7.415	1.975	.144
Group * AgeGroup	2.733	2	1.366	.364	.696
Group * AgeGroup * Age	8.542	2	4.271	1.138	.324
Error	397.943	106	3.754		
Total	83228.312	118			
Corrected Total	2228.550	117			

a. R Squared = .821 (Adjusted R Squared = .803)

Analysis 4

Tests of Between-Subjects Effects

Dependent Variable:Strength (units)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1813.482[a]	7	259.069	68.658	.000
Intercept	2791.621	1	2791.621	739.827	.000
Group	520.962	2	260.481	69.032	.000
AgeGroup	1207.299	1	1207.299	319.955	.000
Age	22.964	1	22.964	6.086	.015
AgeGroup * Age	1328.395	1	1328.395	352.047	.000
Group * AgeGroup	3.861	2	1.930	.512	.601
Error	415.068	110	3.773		
Total	83228.312	118			
Corrected Total	2228.550	117			

a. R Squared = .814 (Adjusted R Squared = .802)

Analysis 5

Tests of Between-Subjects Effects

Dependent Variable:Strength (units)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1809.621[a]	5	361.924	96.760	.000
Intercept	2860.588	1	2860.588	764.775	.000
Group	529.833	2	264.916	70.825	.000
AgeGroup	1215.091	1	1215.091	324.853	.000
Age	22.634	1	22.634	6.051	.015
AgeGroup * Age	1344.445	1	1344.445	359.436	.000
Error	418.928	112	3.740		
Total	83228.312	118			
Corrected Total	2228.550	117			

a. R Squared = .812 (Adjusted R Squared = .804)

Analysis 6

Parameter Estimates

Dependent Variable:Strength (units)

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	46.701	1.743	26.791	.000	43.247	50.155
[Group=1]	5.166	.434	11.902	.000	4.306	6.027
[Group=2]	2.529	.441	5.732	.000	1.655	3.403
[Group=3]	0[a]	.	.	.	.	.
Age	-.443	.032	-13.809	.000	-.507	-.380
[AgeGroup=1]	-39.065	2.167	-18.024	.000	-43.360	-34.771
[AgeGroup=2]	0[a]	.	.	.	.	.
[AgeGroup=1] * Age	1.018	.054	18.959	.000	.912	1.124
[AgeGroup=2] * Age	0[a]	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

Parameter Estimates

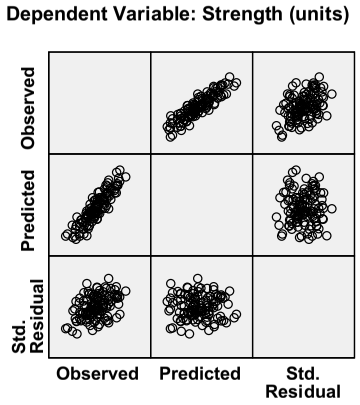


Figure 4

Model: Intercept + Group + AgeGroup + Age + Age

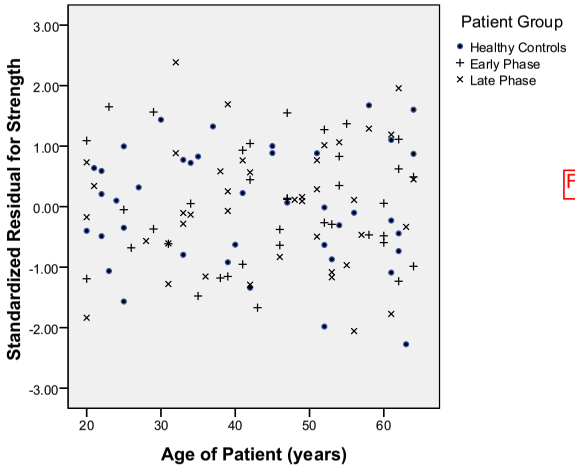


Figure 5

Tests of Between-Subjects Effects

Dependent Variable: Strength (units)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1682.301 [a]	4	420.575	87.003	.000
Intercept	6.040	1	6.040	1.249	.266
Group	533.867	2	266.934	55.219	.000
Age	1098.077	1	1098.077	227.154	.000
AgeSq	1217.692	1	1217.692	251.898	.000
Error	546.249	113	4.834		
Total	83228.312	118			
Corrected Total	2228.550	117			

a. R Squared = .755 (Adjusted R Squared = .746)

Analysis 7

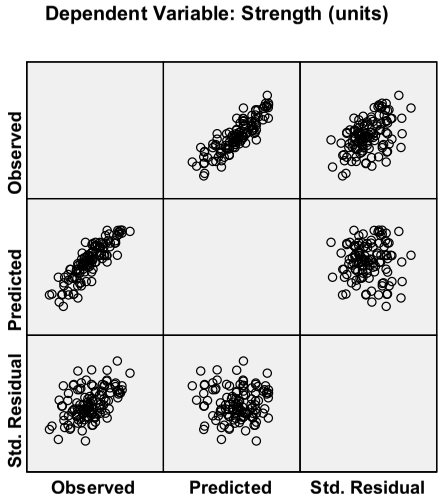


Figure 6

Model: Intercept + Group + Age + AgeSq