Stat 346 Homework 7

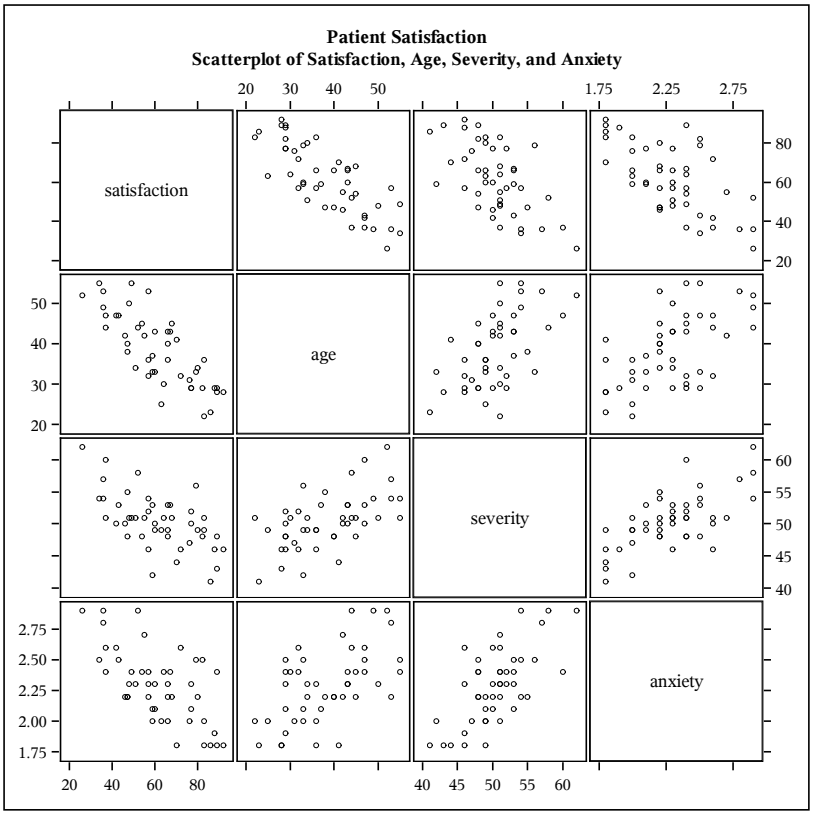
Nathan Jarus

Apr. 10, 2014

1 Problem 1

1.1 Part a

18:56 Wednesday, April 9, 2014 1



Patient Satisfaction

Pairwise correlations

The CORR Procedure

4 Variables:	satisfaction age severity anxiety

Simple Statistics						
Variable	**N**	**Mean**	**Std Dev**	**Sum**	**Minimum**	**Maximum**
satisfaction	46	61.56522	17.23646	2832	26.00000	92.00000
age	46	38.39130	8.91809	1766	22.00000	55.00000
severity	46	50.43478	4.31356	2320	41.00000	62.00000
anxiety	46	2.28696	0.29934	105.20000	1.80000	2.90000

Pearson Correlation Coefficients, N = 46					**Prob > |r| under H0: Rho=0**			
	satisfaction	**age**	**severity**	**anxiety**	**Psatisfaction**	**Page**	**Pseverity**	**Panxiety**
satisfaction	1.00000	−0.78676	−0.60294	−0.64459		<.0001	<.0001	<.0001
age	−0.78676	1.00000	0.56795	0.56968	<.0001		<.0001	<.0001
severity	−0.60294	0.56795	1.00000	0.67053	<.0001	<.0001		<.0001
anxiety	−0.64459	0.56968	0.67053	1.00000	<.0001	<.0001	<.0001	

There seems to be a pretty strong linear relationship between satisfaction and age, severity, and anxiety.

1.2 Part b

Patient Satisfaction

MLR with Age and Anxiety

The REG Procedure

Model: MODEL1

Dependent Variable: satisfaction

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	2	9038.80461	4519.40231	44.88	<.0001
Error	43	4330.49973	100.70930		
Corrected Total	45	13369			

Patient Satisfaction

MLR with Age, severity, and Anxiety

The REG Procedure

Model: MODEL1

Dependent Variable: satisfaction

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	3	9120.46367	3040.15456	30.05	<.0001
Error	42	4248.84068	101.16287		
Corrected Total	45	13369			

$ESS = 4330.49973 - 4248.84068 = 81.65905$
$H_0 : \beta_{severity} = 0$
$H_a : \beta_{severity} \neq 0$
In this case, we have 46 observations. As such, we have 1 and 42 degrees of freedom for the numerator and denominator.
Our F-statistic is $F = \frac{81.65905}{101.1627} = 0.8072$
The p-value for $F(0.95; 1, 42) \approx 4.0487$, meaning that we fail to reject the null hypothesis.

1.3 Part b

Patient Satisfaction

General Linear Test with Age, Severity, and Anxiety

The REG Procedure

Model: MODEL1

Test TEST1 Results for Dependent Variable satisfaction				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	81.65905	0.81	0.3741
Denominator	42	101.16287		

Again, we fail to reject the null, meaning that we cannot say that severity has a significant linear relationship to satisfaction.

1.4 Part c

Patient Satisfaction

Null Hypothesis Test for Severity and Anxiety

The REG Procedure

Model: MODEL1

Test TEST1 Results for Dependent Variable satisfaction				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	2	422.53741	4.18	0.0222
Denominator	42	101.16287		

$H_0 : \beta_{severity} = \beta_{anxiety} = 0$
$H_a : \beta_{severity} \neq 0$ and/or $\beta_{anxiety} \neq 0$
Full model: $Y = \beta_{severity}X_{severity} + \beta_{anxiety}X_{anxiety} + \beta_{age}X_{age} + \beta_0$
Reduced model: $Y = \beta_{age}X_{age} + \beta_0$
$df_F = 42$
$df_R = 44$
Test statistic: $\frac{(SSE(R)-SSE(F))/(df_R-df_F)}{SSE(F)/df_F}$
P value: $F(0.95; 2, 44) = 3.2317$.
Conclusion: Reject H_0; at least one of severity and anxiety have a significant linear relationship with satisfaction.

2 Problem 2

Patient Satisfaction

Predictors for MLR

The REG Procedure

Model: MODEL1

Dependent Variable: satisfaction

Number of Observations Read	46
Number of Observations Used	46

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	3	9120.46367	3040.15456	30.05	<.0001
Error	42	4248.84068	101.16287		
Corrected Total	45	13369			

Root MSE	10.05798	**R-Square**	0.6822
Dependent Mean	61.56522	**Adj R-Sq**	0.6595
Coeff Var	16.33711		

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	158.49125	18.12589	8.74	<.0001
age	1	−1.14161	0.21480	−5.31	<.0001
severity	1	−0.44200	0.49197	−0.90	0.3741
anxiety	1	−13.47016	7.09966	−1.90	0.0647

Parameter Estimates Continued					
Variable	**Type I SS**	**Type II SS**	**Squared Partial Corr Type I**	**Squared Partial Corr Type II**	**Variance Inflation**
Intercept	174353	7734.51573	.	.	0
age	8275.38885	2857.55338	0.61898	0.40211	1.63230
severity	480.91529	81.65905	0.09441	0.01886	2.00324
anxiety	364.15952	364.15952	0.07894	0.07894	2.00906

2.1 Part a

All the VIFs are smaller than 10, so multicollinearity is not an issue here.

2.2 Part b

Sum of Type I sums of squares: 9120.46366
Sum of Type II/III sums of squares: 3303.37249
The Type I sums of squares is the same as the model sums of squares, as expected.
For anxiety, the Type I and Type II/III sums of squares are the same because it was added last, so the sequential sums of squares (Type I) is measuring the same thing as the variable-added-last sums of squares (Type III).

2.3 Part c

We use Type II sums of squares here because we are concerned about how much the variable explains after the other variables are included in the model. Age explains the most after the other two variables are taken into account: 40%.
Neither Type I nor Type II partial determination coefficients sum to R^2.

2.4 Part d

$R^2_{2|1} = \frac{480.91529}{13369 - 8275.38885} = 0.9441$

3 Problem 3

3.1 Part a

Patient Satisfaction

General Linear Model

The GLM Procedure

Number of Observations Read	46
Number of Observations Used	46

Patient Satisfaction

General Linear Model

The GLM Procedure

Dependent Variable: satisfaction

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	9038.80461	4519.40231	44.88	<.0001
Error	43	4330.49973	100.70930		
Corrected Total	45	13369.30435			

R-Square	Coeff Var	Root MSE	satisfaction Mean
0.676086	16.30044	10.03540	61.56522

Source	DF	Type I SS	Mean Square	F Value	Pr > F
age	1	8275.388851	8275.388851	82.17	<.0001
anxiety	1	763.415762	763.415762	7.58	0.0086

Source	DF	Type III SS	Mean Square	F Value	Pr > F
age	1	3483.891467	3483.891467	34.59	<.0001
anxiety	1	763.415762	763.415762	7.58	0.0086

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	145.9412281	11.52509259	12.66	<.0001
age	−1.2004715	0.20410532	−5.88	<.0001
anxiety	−16.7420519	6.08083056	−2.75	0.0086

Since the p-value for the test, 0.0086, is less than 0.05, we reject the null and conclude that anxiety does have a significant linear relationship with satisfaction in a model containing age.

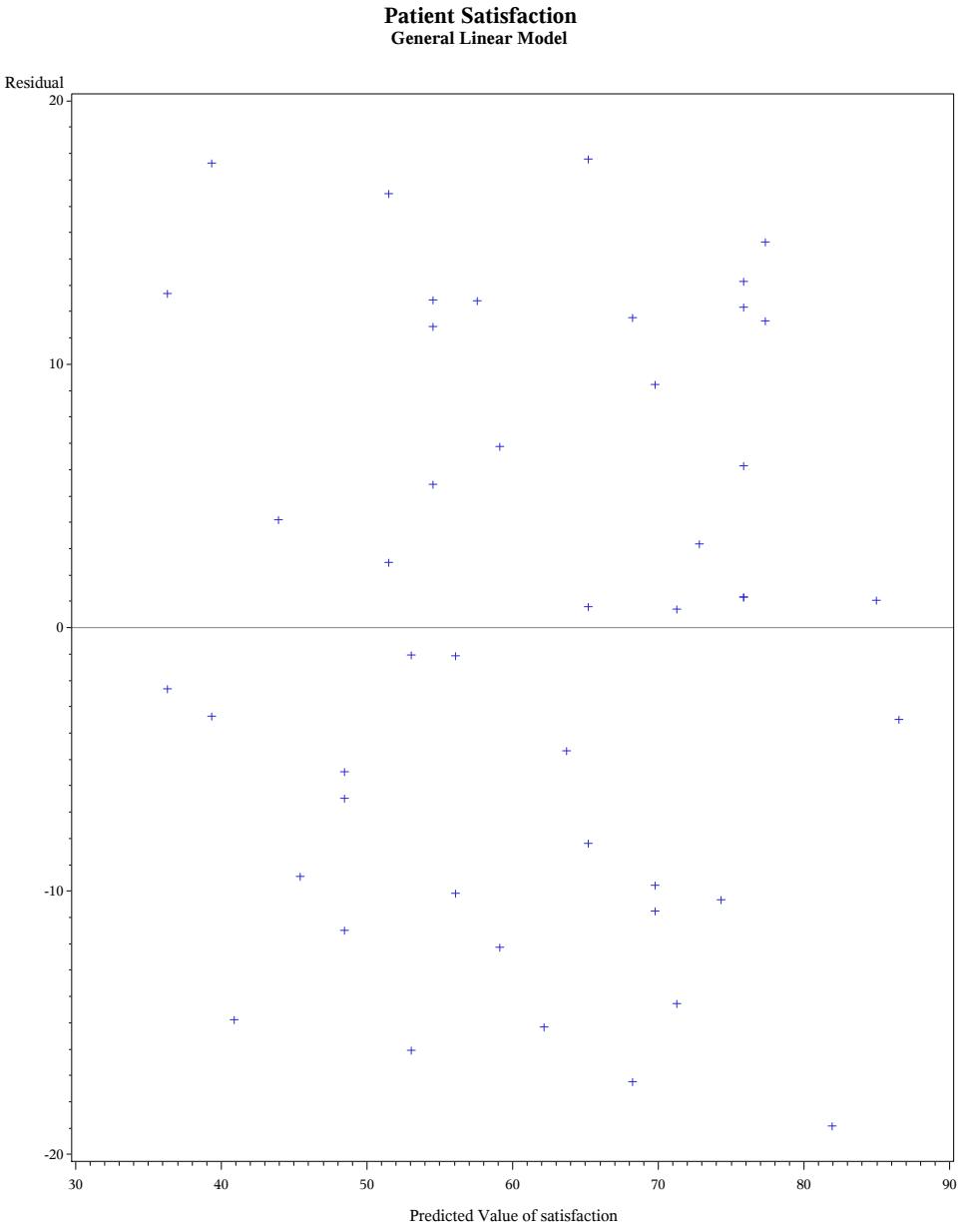
3.2 Part b

Model	Adjusted R^2
Age	0.6103
Age, Anxiety	0.6610
Age, Severity, Anxiety	0.6595

Age and Anxiety have the highest R^2 value, which makes sense given that we could not conclude a significant linear relationship between severity and satisfaction.

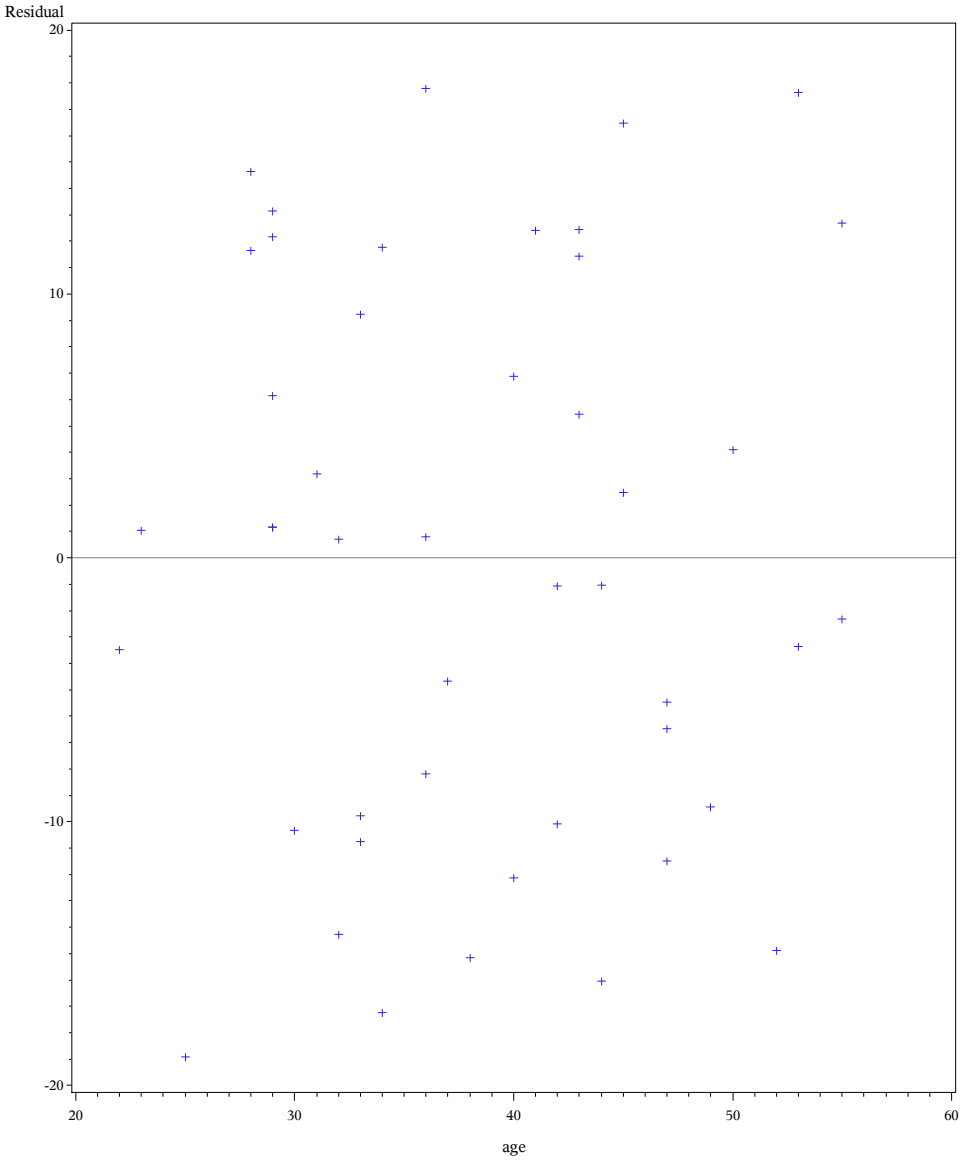
3.3 Part c

3.3.1 Model: Age

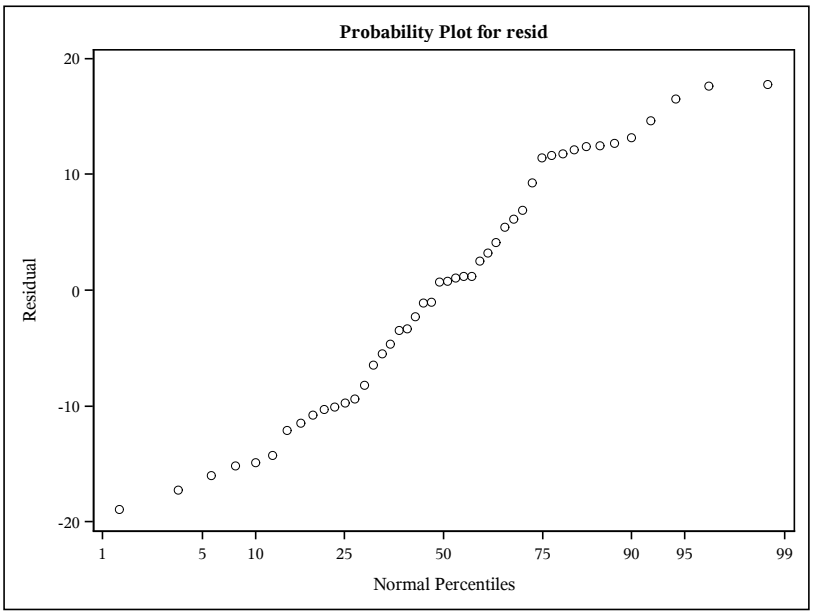


18:56 Wednesday, April 9, 2014 **2**

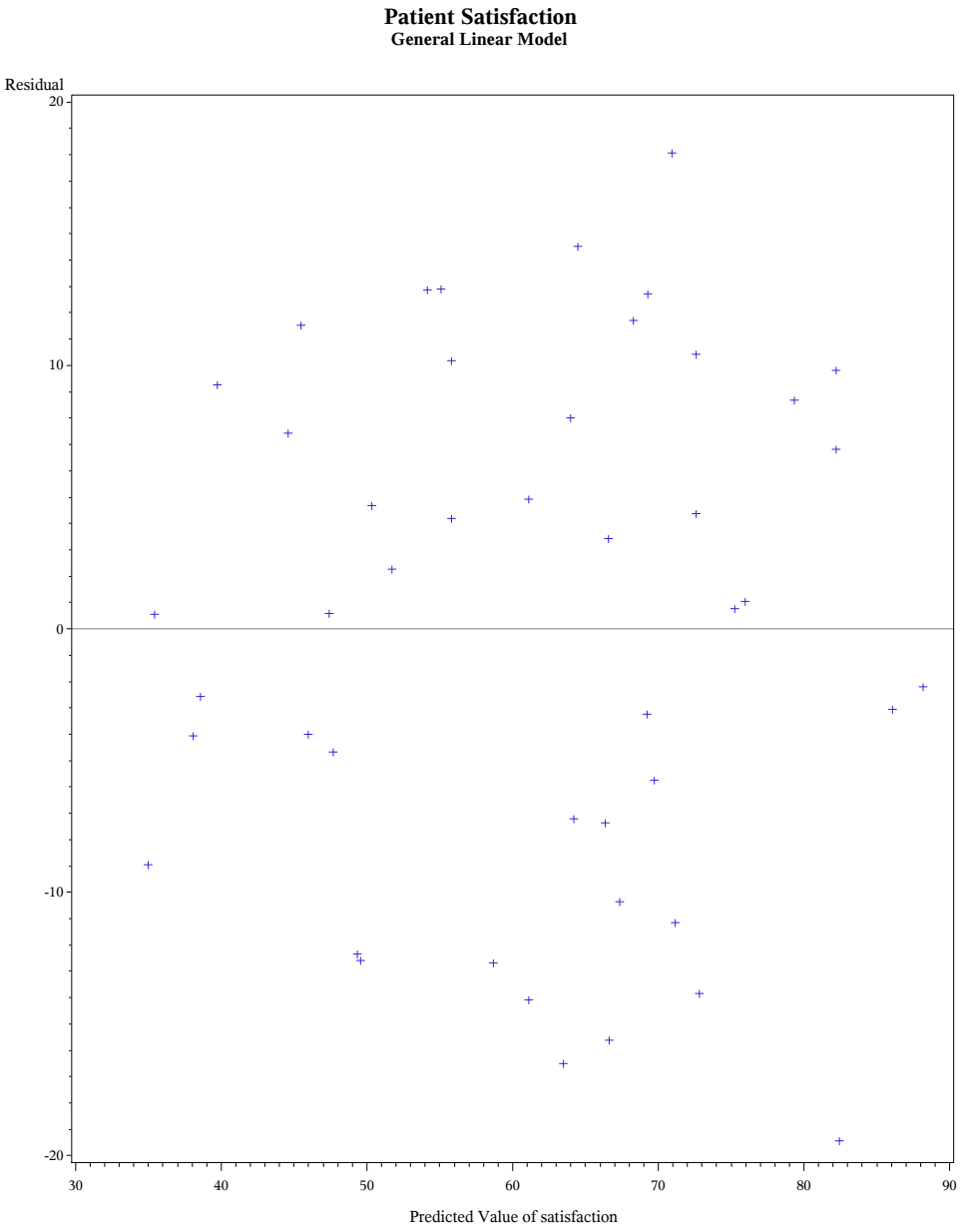
Patient Satisfaction
General Linear Model



Patient Satisfaction 18:56 Wednesday, April 9, 2014 **1**
General Linear Model

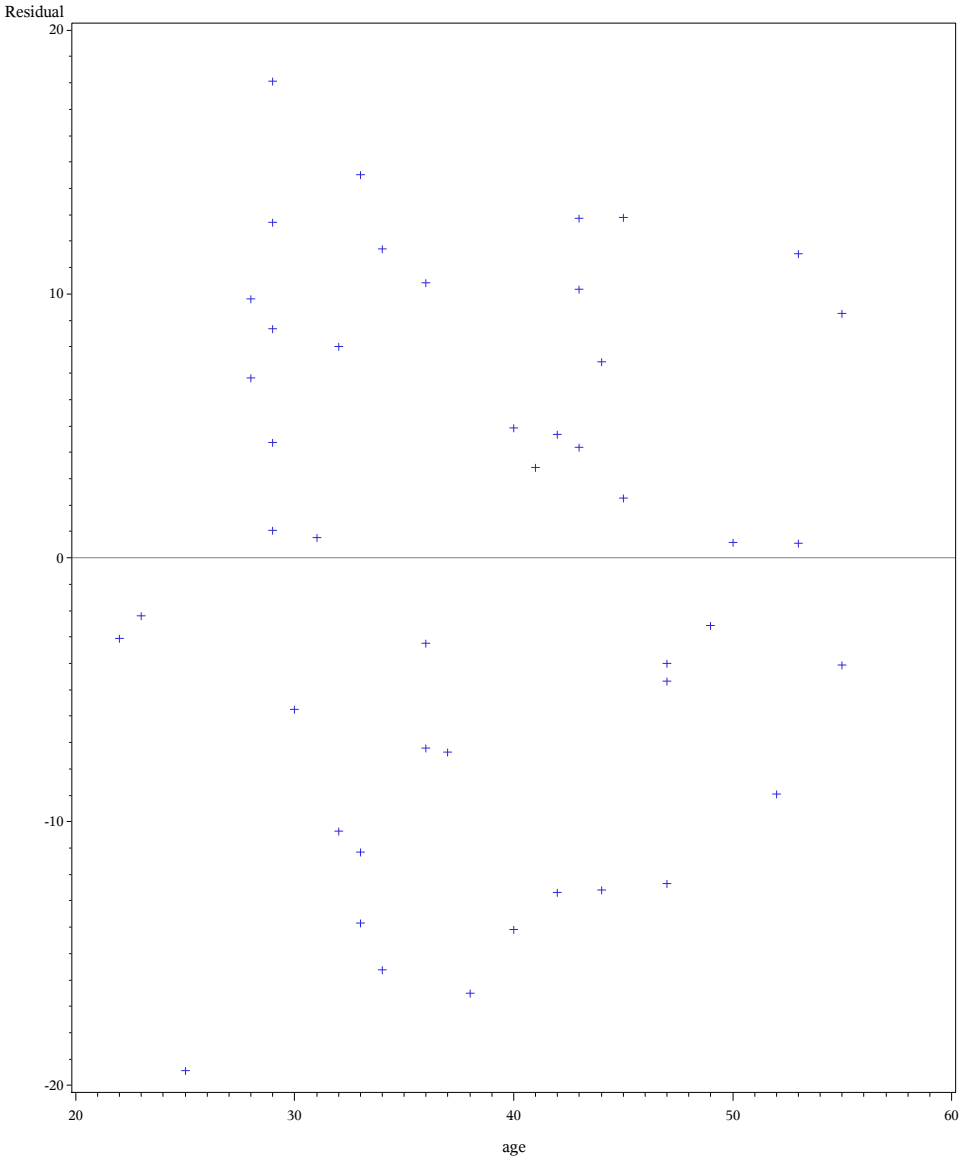


3.3.2 Model: Age, Anxiety



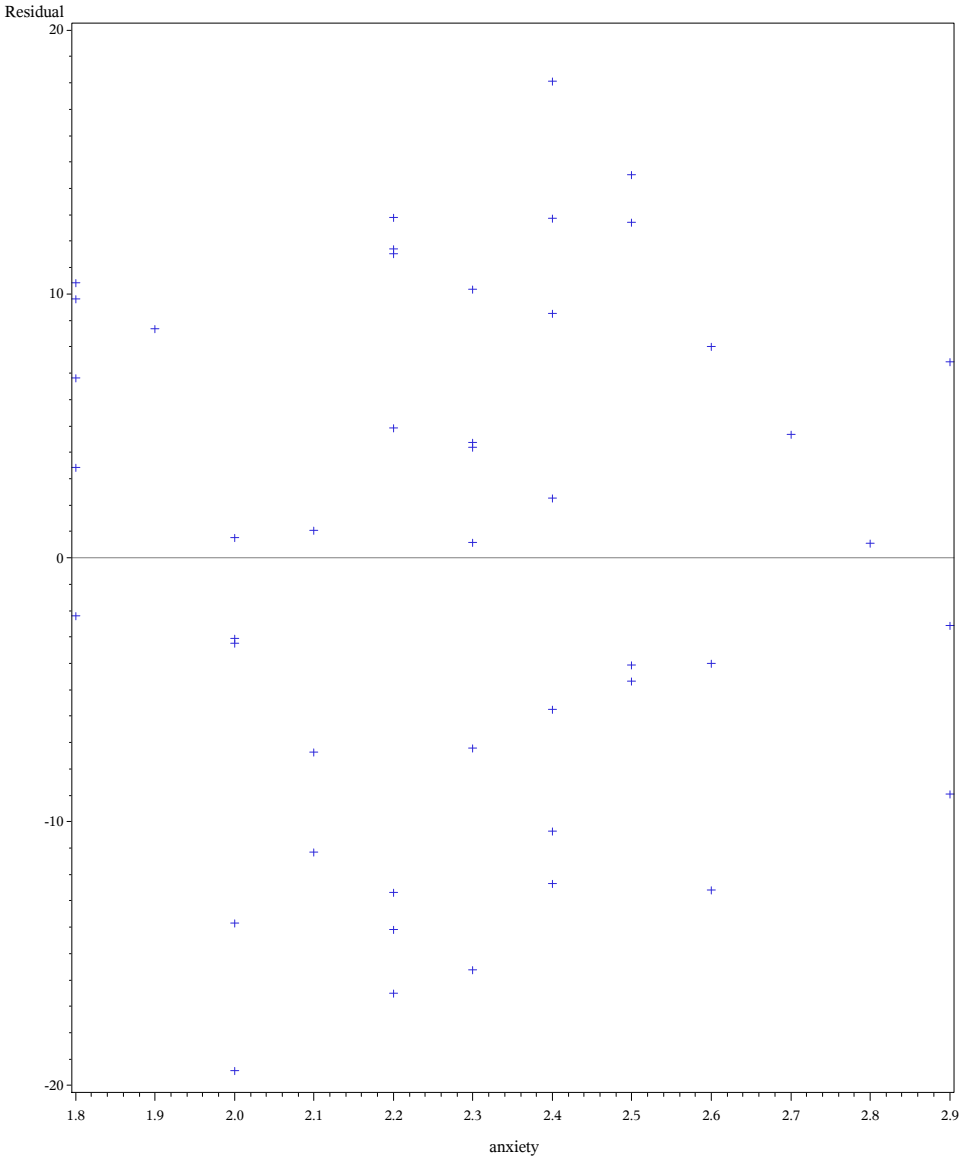
18:56 Wednesday, April 9, 2014 **2**

Patient Satisfaction
General Linear Model

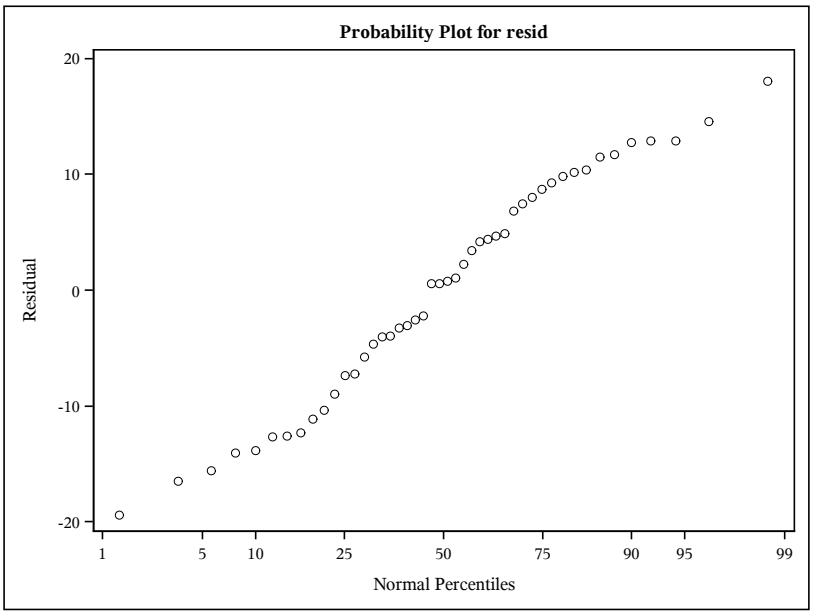


18:56 Wednesday, April 9, 2014 3

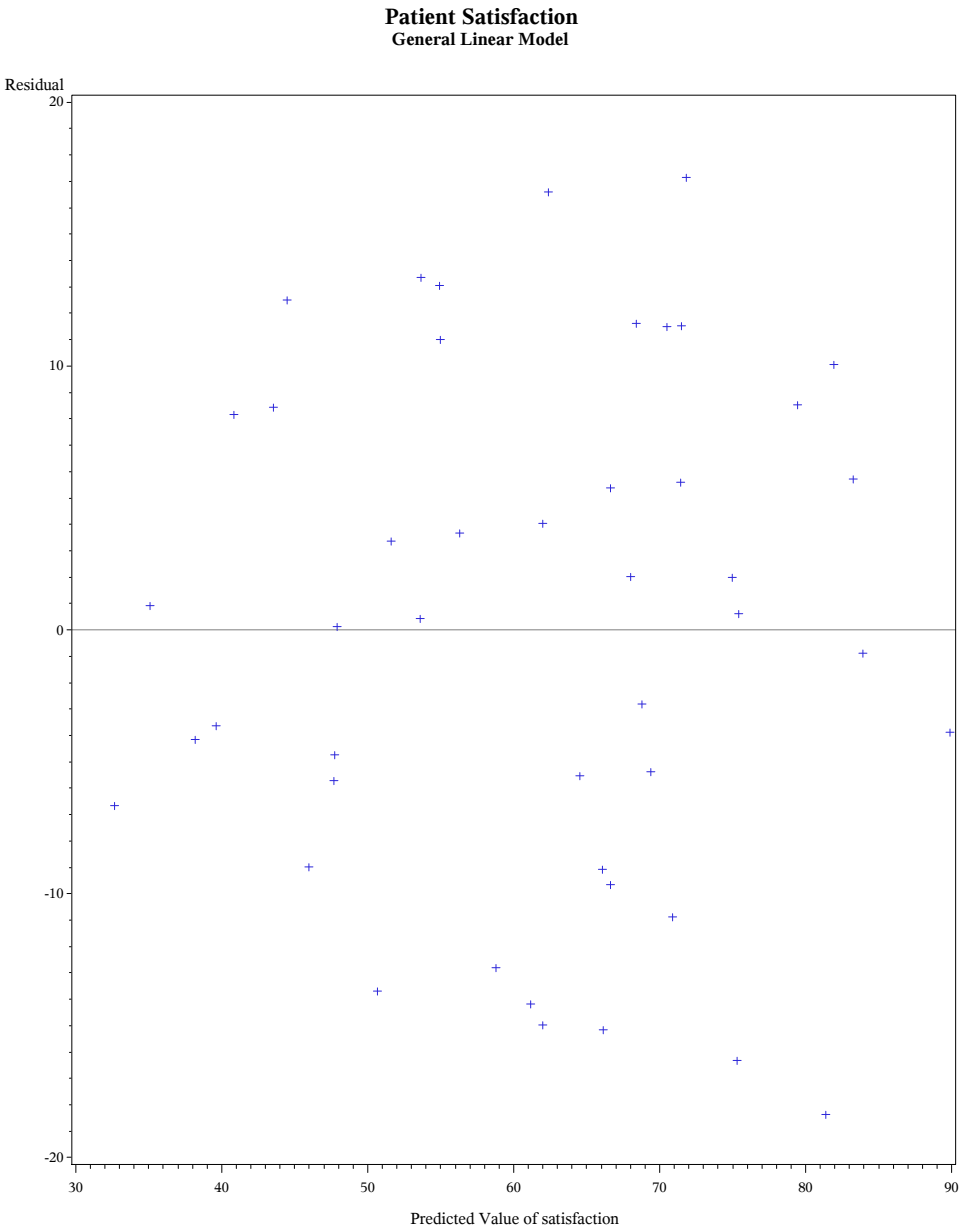
Patient Satisfaction
General Linear Model



Patient Satisfaction 18:56 Wednesday, April 9, 2014 **1**
General Linear Model

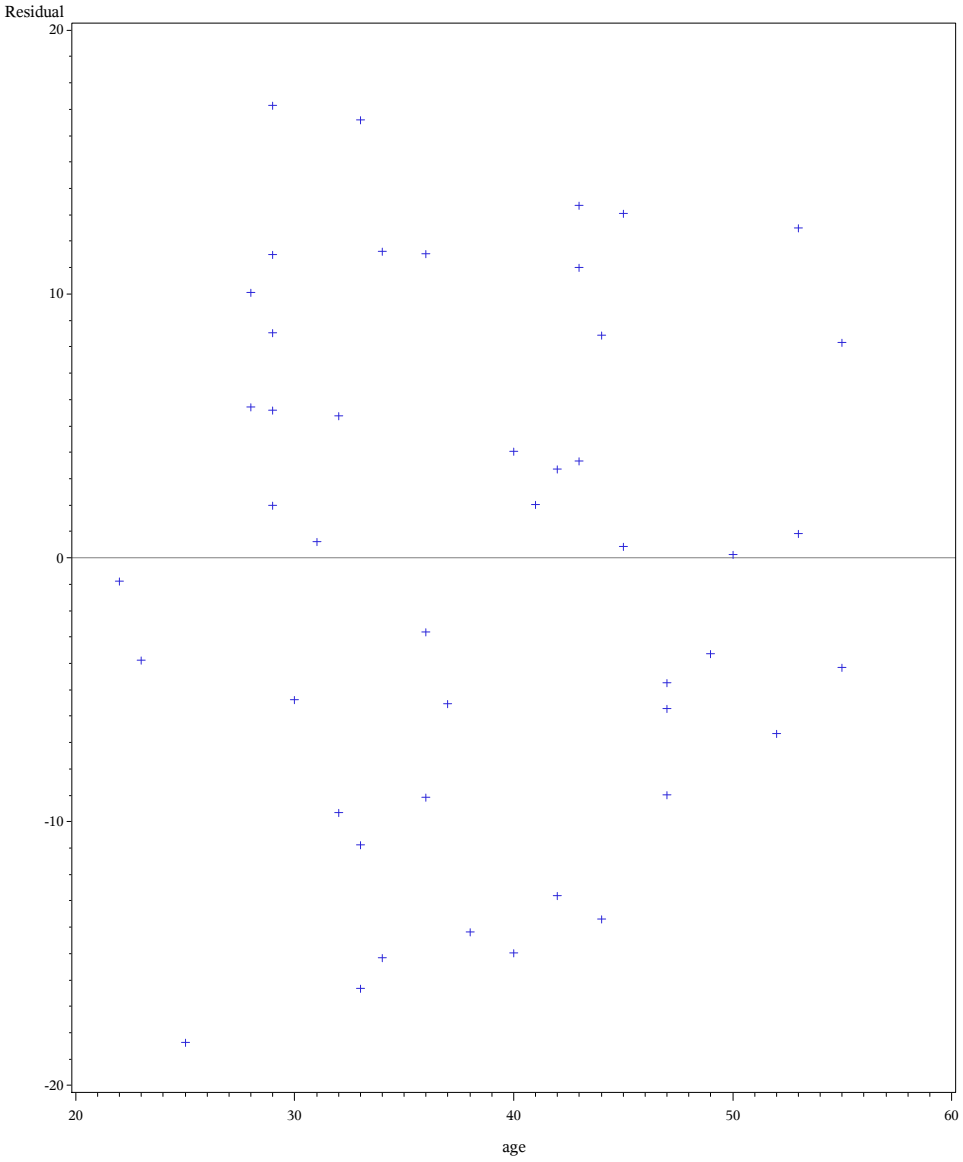


3.3.3 Model: Age, Severity, Anxiety



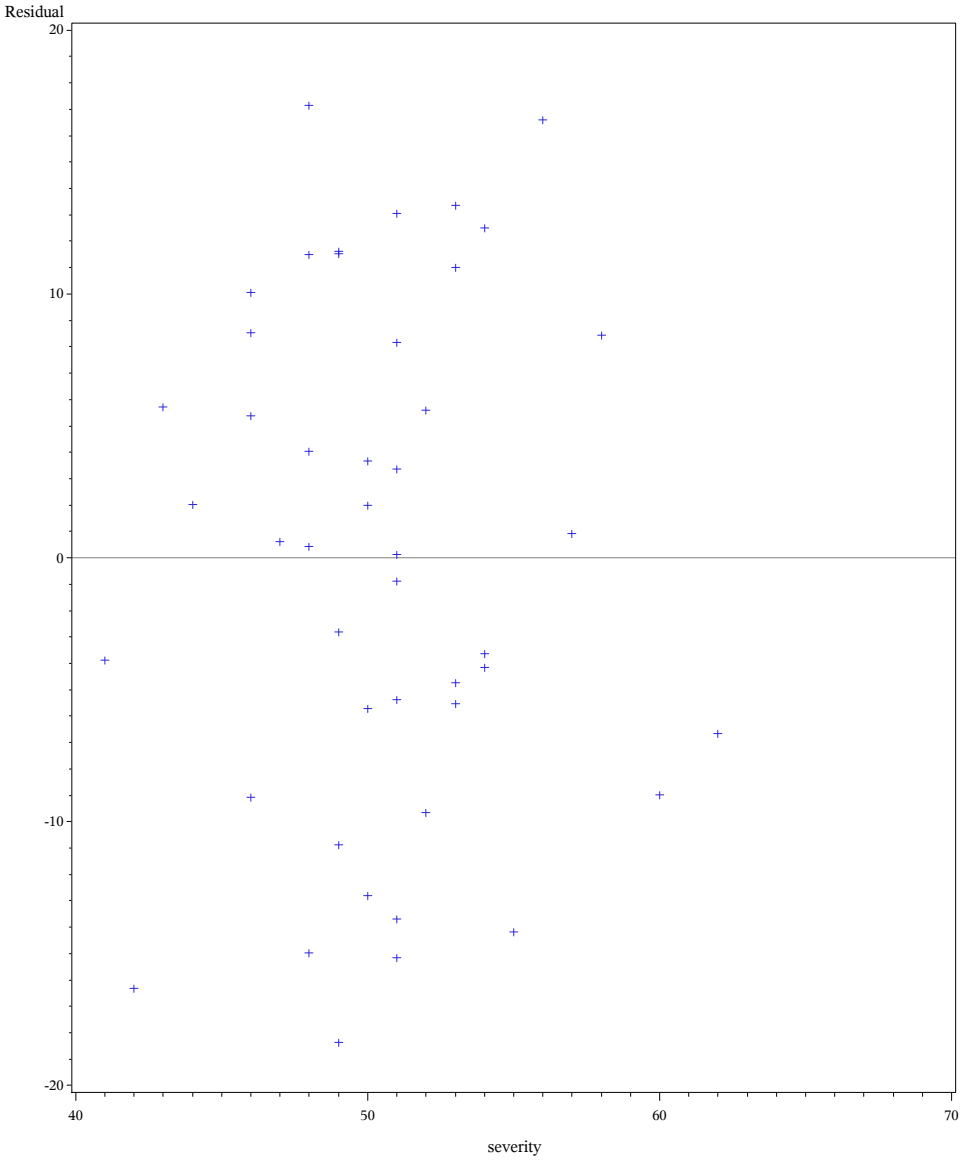
18:56 Wednesday, April 9, 2014 **2**

Patient Satisfaction
General Linear Model



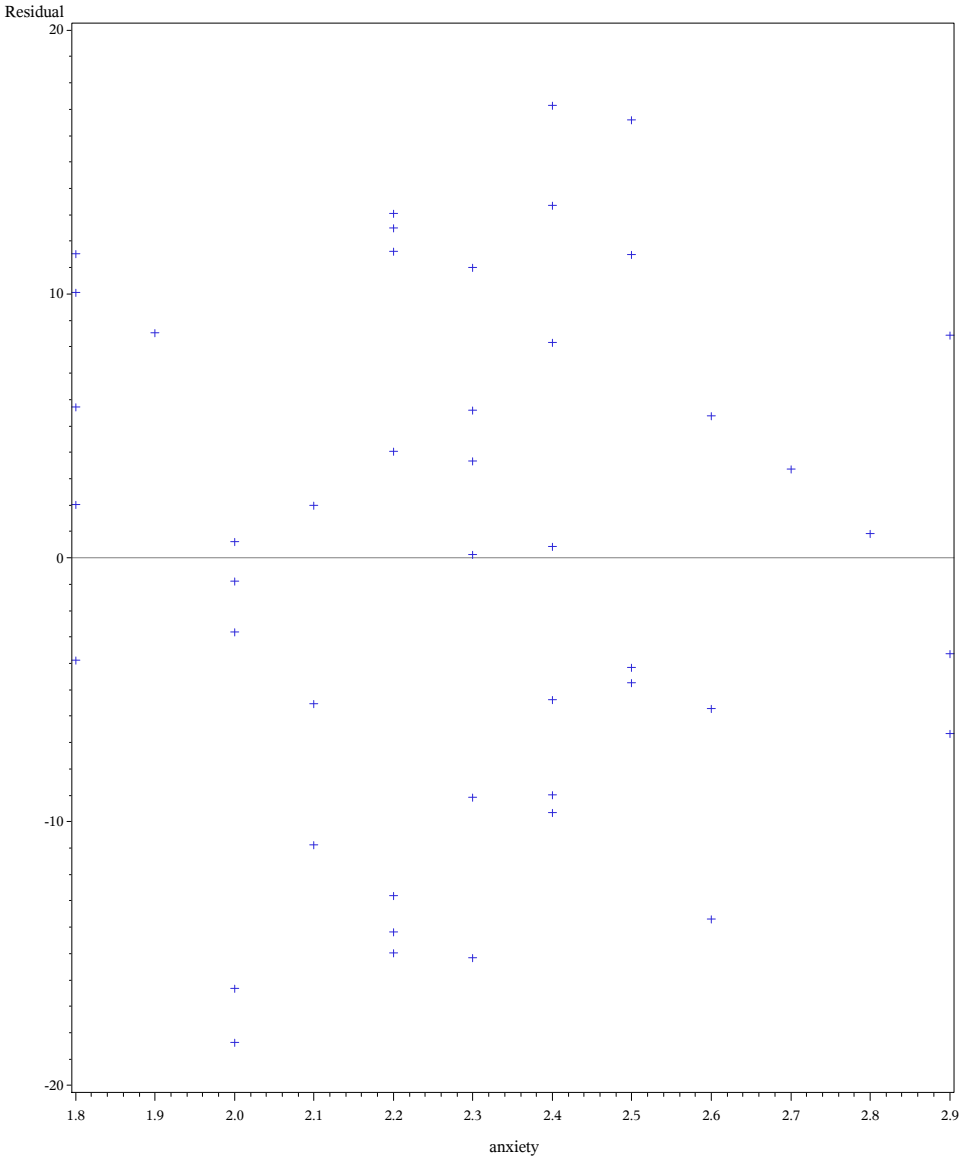
18:56 Wednesday, April 9, 2014 **3**

Patient Satisfaction
General Linear Model

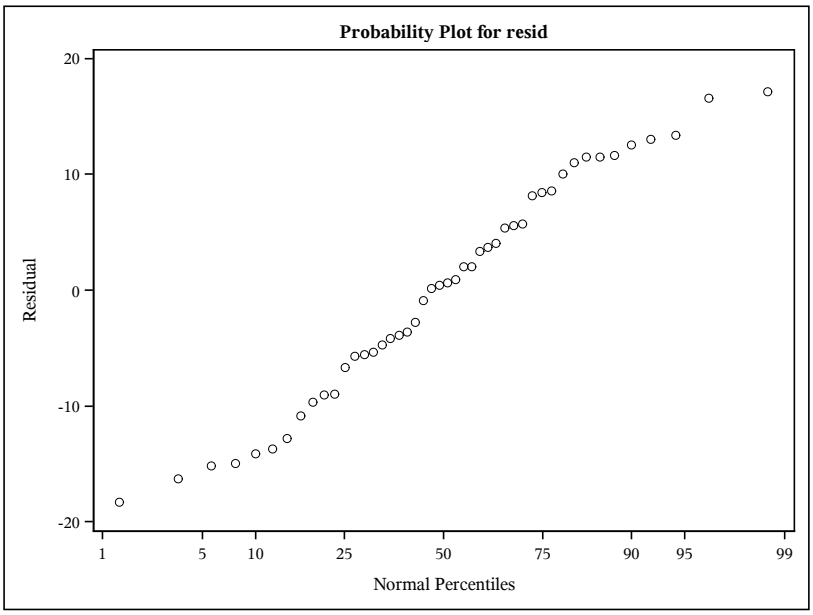


18:56 Wednesday, April 9, 2014 **4**

Patient Satisfaction
General Linear Model



Patient Satisfaction 18:56 Wednesday, April 9, 2014 **1**
General Linear Model



None of the plots indicate violations of MLR assumptions.

4 Problem 4

4.1 Part 1

- CT Ortman
- Sarah Katterhenry
- Ryan Schmitt
- Nathan Jarus

4.2 Part 2

We will be using a dataset of diabetes data.
Description: `http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/diabetes.html`
Variable information: `http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/Cdiabetes.html`
Dataset: `http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/diabetes.xls`

Variable	Description	Type
id	Subject ID	quantitative
chol	Total Cholesterol	quantitative
stab.glu	Stabilized Glucose	quantitative
hdl	High Density Lipoprotein	quantitative
ratio	Cholesterol/HDL Ratio	quantitative
glyhb	Glycosolated Hemoglobin	quantitative
location	Louisia or Buckingham	categorical
age	years	quantitative
gender	male/female	categorical
height	inches	quantitative
weight	pounds	quantitative
frame	small/med/large	categorical
bp.1s	First Systolic Blood Pressure	quantitative
bp.1d	First Diastolic Blood Pressure	quantitative
bp.2s	Second Systolic Blood Pressure	quantitative
bp.2d	Second Diastolic Blood Pressure	quantitative
waist	inches	quantitative
hip	inches	quantitative
time.ppn	Postprandial Time when Labs were Drawn	quantitative

4.3 Part 3

We will attempt to predict Glycosolated Hemoglobin from other variables (excluding id) as this is used as an positive diagnosis of diabetes.
We will also attempt to predict Stabilized Glucose from other variables (again, excluding id) as this indicates how well the patient is managing their diabetes.

4.4 Part 4

Each predicted variable will be modeled with multiple linear regression on all variables. Diagnostics will be performed to determine which variables provide the strongest predictors. Any actions suggested by the diagnostic results will be performed, and a final model will be developed.