Exercises unit 1.4: Multiple Regression

1) Obtaining the regression model

Table of analysis variance (with CoastDistance)

		Standard	T	
Parameter	*Estimate*	*Error*	*Statistic*	*P-Value*
CONSTANT	2900,65	543,153	5,3404	0,0000
xcoord(east)	0,00222026	0,000285089	7,78795	0,0000
ycoord(north)	-0,000987919	0,000167087	-5,91261	0,0000
height	0,0638771	0,00849455	7,51978	0,0000
Coastdistance	0,000483279	0,000339538	1,42334	0,1557

Regression model is explanatory → We can use this model in order to explain Y

Analysis of Variance

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Model	131955,	4	32988,6	51,07	0,0000
Residual	187330,	290	645,966		
Total (Corr.)	319285,	294			

Standard Error of Est. = 25,4159

T-statistic table

P-value is bigger than 0,5, so in this case can be Xj eliminated from the model (it is not explanatory variable) → CoastDistance is deleted and

EQUATION OF MODEL

factorR = 2388,72 + 0,00187994*xcoord(east) - 0,000814253*ycoord(north) + 0,0660843*height

Table of analysis variance (without CoastDistance)

		Standard	T	
Parameter	*Estimate*	*Error*	*Statistic*	*P-Value*
CONSTANT	2388,72	407,72	5,85873	0,0000
xcoord(east)	0,00187994	0,000155531	12,0872	0,0000
ycoord(north)	-0,000814253	0,000114354	-7,12046	0,0000
height	0,0660843	0,00836653	7,89865	0,0000

Analysis of Variance

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Model	130646,	3	43548,6	67,18	0,0000
Residual	188639,	291	648,243		
Total (Corr.)	319285,	294			

Standard Error of Est. = 25,4606

2) Hypothesis of the regression model

Tests for Normality for Data_FactorR.RESIDUALS

Test	*Statistic*	*P-Value*
Chi-Square	39,7559	0,229049
Shapiro-Wilk W	0,97268	0,0100609
Skewness Z-score	1,80232	0,0714951
Kurtosis Z-score	-0,0646	0,948487

In case, that p-value of residuals (errors) is smaller than 0,05 → data don't follow normal distribution. *(Save residuals → Describe → Fitting unfitted data)*

Goodness-of-Fit Tests for Data_FactorR.RESIDUALS

Kolmogorov-Smirnov Test

	Normal
DPLUS	0,0775578
DMINUS	0,0351592
DN	0,0775578
P-Value	0,0575089

Kuiper V

	Normal
V	0,112717
Modified Form	1,95502
P-Value	<0.05

Watson U^2

	Normal
U^2	0,189968
Modified Form	0,190144
P-Value	<0.05

Modified Kolmogorov-Smirnov D

	Normal
D	0,0775578
Modified Form	1,3419
P-Value	<0.10

Cramer-Von Mises W^2

	Normal
W^2	0,22642
Modified Form	0,225834
P-Value	>=0.10

Anderson-Darling A^2

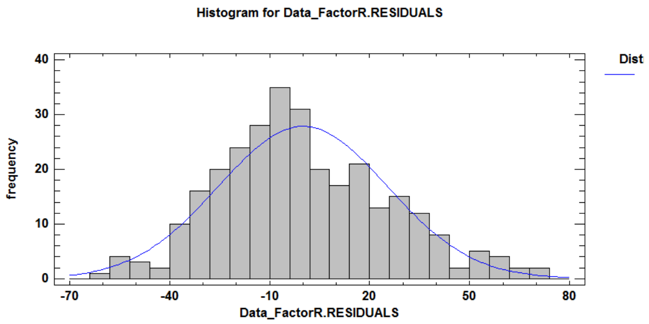
	Normal
A^2	1,25948
Modified Form	1,25948
P-Value	>=0.10

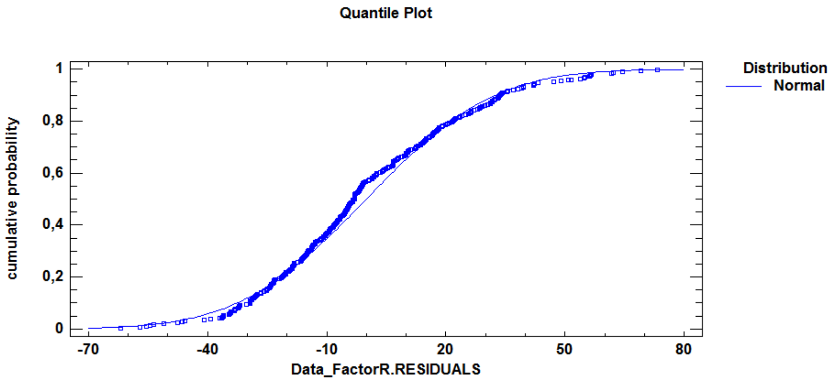
Chi-Square Test

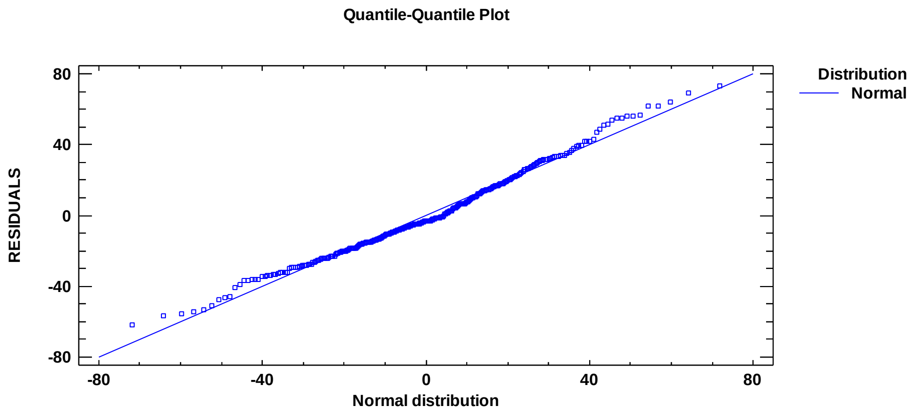
	Lower Limit	*Upper Limit*	*Observed Frequency*	*Expected Frequency*	*Chi-Square*
at or below		-48,7966	6	7,97	0,49
	-48,7966	-40,6998	4	7,97	1,98
	-40,6998	-35,4078	6	7,97	0,49
	-35,4078	-31,325	11	7,97	1,15
	-31,325	-27,9253	10	7,97	0,52
	-27,9253	-24,966	8	7,97	0,00
	-24,966	-22,3133	11	7,97	1,15
	-22,3133	-19,8853	9	7,97	0,13
	-19,8853	-17,6277	10	7,97	0,52
	-17,6277	-15,5021	7	7,97	0,12
	-15,5021	-13,4806	13	7,97	3,17
	-13,4806	-11,5416	7	7,97	0,12
	-11,5416	-9,66801	7	7,97	0,12
	-9,66801	-7,84594	10	7,97	0,52
	-7,84594	-6,06364	9	7,97	0,13
	-6,06364	-4,31089	13	7,97	3,17
	-4,31089	-2,57856	13	7,97	3,17
	-2,57856	-0,858227	12	7,97	2,03
	-0,858227	0,858226	4	7,97	1,98
	0,858226	2,57856	5	7,97	1,11
	2,57856	4,31088	5	7,97	1,11
	4,31088	6,06363	5	7,97	1,11
	6,06363	7,84594	8	7,97	0,00
	7,84594	9,66801	4	7,97	1,98
	9,66801	11,5416	7	7,97	0,12
	11,5416	13,4806	4	7,97	1,98
	13,4806	15,5021	9	7,97	0,13
	15,5021	17,6277	7	7,97	0,12
	17,6277	19,8853	8	7,97	0,00
	19,8853	22,3133	7	7,97	0,12
	22,3133	24,966	5	7,97	1,11
	24,966	27,9253	6	7,97	0,49
	27,9253	31,325	7	7,97	0,12
	31,325	35,4078	13	7,97	3,17
	35,4078	40,6998	6	7,97	0,49
	40,6998	48,7966	5	7,97	1,11
above	48,7966		14	7,97	4,56

Chi-Square = 39,7559 with 34 d.f.   P-Value = 0,229049

Chi-Squared Test, Watson and Kupier rejected normal distribution. So in this case is obligatory to watch histogram, curtosis and so on; in order to decide if data follow normal distribution or not.





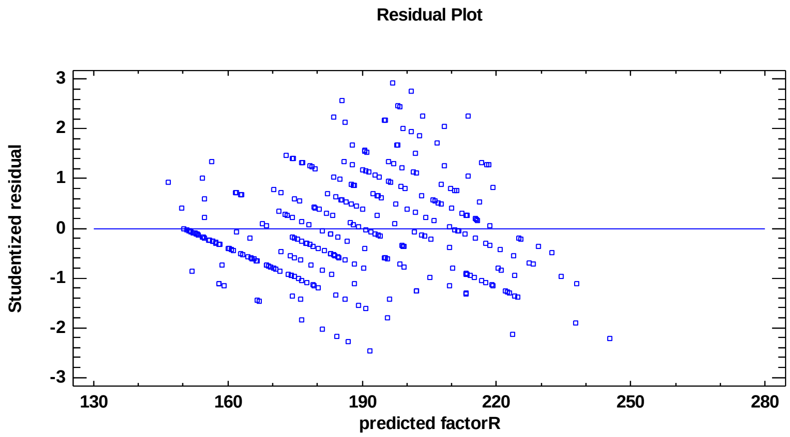


Data are separated (down and up) by the line.

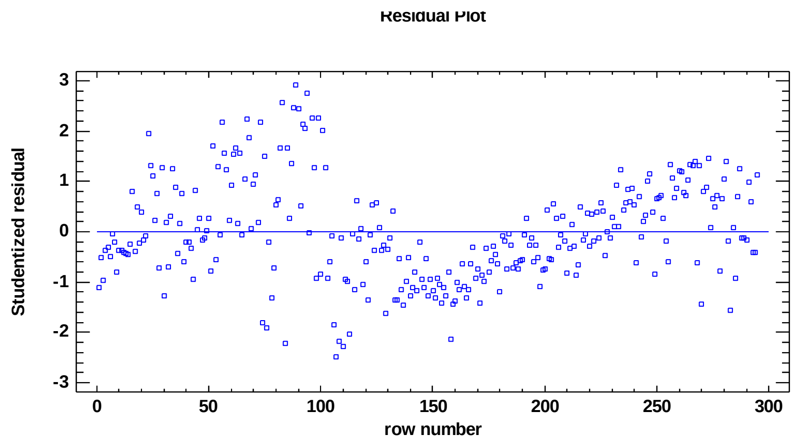
Plot follow a normal distribution, in the middle of the plot there are some deviations.

Here we observe data that are separated from a straight line (at the beginning and at the end)

We cannot reject the normality idea because the most of the tests indicate it. When we observe histogram and plots we cannot reject idea that data follow normal distribution.



Variation is not constant, because FactorR, based on previous analysis, is higher. Doesn't folllow homosedascity hypotesis, so we can say, that linear hypotesis was fulfilled. (*Tables and Graphs → Residuals versus Row Number*)



We observe a tendence in residuals. First residuals are positive, then they are negative and at the end we have positive and negative. Residuals depend from row number. This is a relationship between row number and residuals. If exists relationship we can conclude that independence of residuals is not fulfilled.

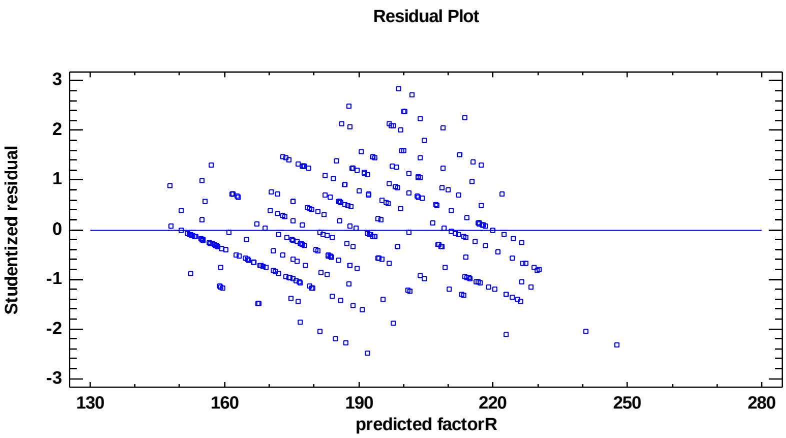
The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the P-value is less than 0,05, there is an indication of possible serial correlation at the 95,0% confidence level. Plot the residuals versus row order to see if there is any pattern that can be seen. Lag 1 residual autocorrelation = 0,396113

If this parameter is near to 0, we observe correlation between residuals and row number. In this case it is not fulfilled.

Unusual Residuals

		Predicted		*Studentized*
Row	*Y*	*Y*	*Residual*	*Residual*
56	250,0	195,037	54,9628	2,18
67	240,0	183,746	56,2537	2,23
73	250,0	195,223	54,7771	2,17
83	250,0	185,595	64,4051	2,56
84	190,0	245,36	-55,3603	-2,22
88	260,0	197,944	62,0565	2,46
89	270,0	196,824	73,1758	2,92
90	260,0	198,351	61,6493	2,45
92	240,0	186,09	53,9098	2,14
93	260,0	208,416	51,584	2,06
94	270,0	200,966	69,034	2,75
96	270,0	213,645	56,3545	2,26
99	260,0	203,405	56,5953	2,25
101	250,0	199,104	50,8956	2,02
107	130,0	191,795	-61,7949	-2,48
108	130,0	184,44	-54,4404	-2,17
110	130,0	186,935	-56,935	-2,27
113	130,0	181,114	-51,1137	-2,03
158	170,0	223,578	-53,5784	-2,14

The table of unusual residuals lists all observations which have Studentized residuals greater than 2 in absolute value. Studentized residuals measure how many standard deviations each observed value of factorR deviates from a model fitted using all of the data except that observation. In this case, there are 19 Studentized residuals greater than 2, but none greater than 3.



Box-Cox Optimization (Constant in Model)

Dependent variable: factorR

Independent variables: xcoord(east), ycoord(north), height, Coastdistance

Box-Cox transformation applied: power = -0,18125 shift = 0

Cochrane-Orcutt transformation applied: autocorrelation = 0,466825

		Standard	*T*	
Parameter	*Estimate*	*Error*	*Statistic*	*P-Value*
CONSTANT	2375,81	762,244	3,11685	0,0020
xcoord(east)	0,00100702	0,000390697	2,5775	0,0104
ycoord(north)	-0,000339064	0,000231585	-1,4641	0,1443
height	0,0575771	0,00728023	7,90869	0,0000
Coastdistance	-0,000922822	0,000453656	-2,03419	0,0428

Model after using Box-Cox transformation and Cochrane-Orcutt transformation.

P-value of ycoord is higher than 0,05. In this case this variable can be remove from the model. Model has changed

Analysis of Variance

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Model	115454,	4	28863,4	56,84	0,0000
Residual	146765,	289	507,837		
Total (Corr.)	262219,	293			

In determining whether the model can be simplified, notice that the highest P-value on the independent variables is 0,1486, belonging to ycoord(north). Since the P-value is greater or equal to 0,05, that term is not statistically significant at the 95,0% or higher confidence level. Consequently, you should consider removing ycoord(north) from the model.

Ycoord was removed from the model →

Dependent variable: factorR
Independent variables: xcoord(east), height, Coastdistance

Box-Cox transformation applied: power = -0,18125 shift = 0
Cochrane-Orcutt transformation applied: autocorrelation = 0,512871

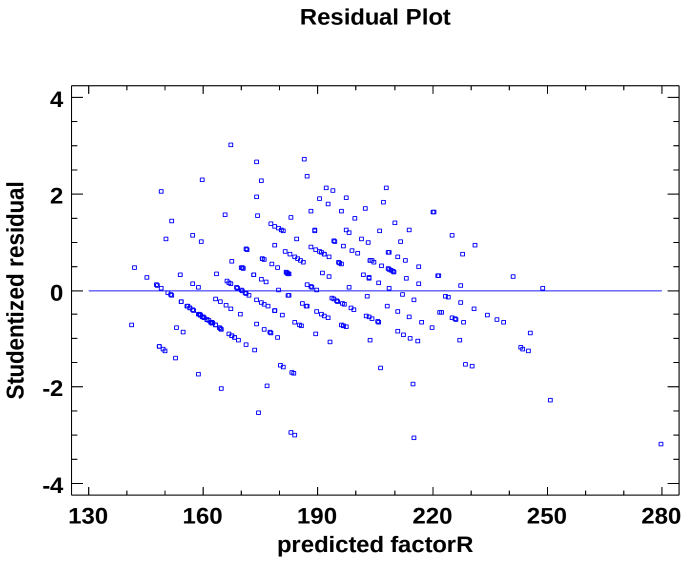
		Standard	*T*	
Parameter	*Estimate*	*Error*	*Statistic*	*P-Value*
CONSTANT	1293,05	123,188	10,4966	0,0000
xcoord(east)	0,000451122	0,000161475	2,79376	0,0056
height	0,0544174	0,00676039	8,04945	0,0000
Coastdistance	-0,0015277	0,000246252	-6,20382	0,0000

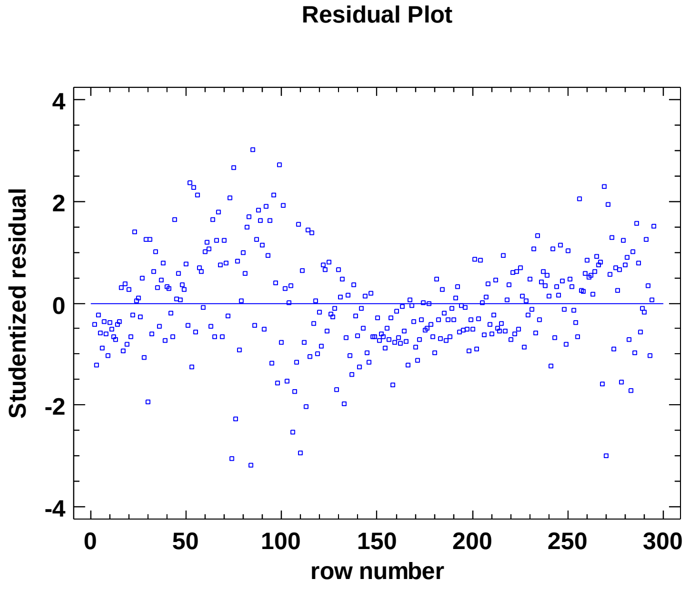
Analysis of Variance

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Model	117048,	3	39016,2	76,76	0,0000
Residual	147394,	290	508,255		
Total (Corr.)	264442,	293			

EQUATION

BoxCox(factorR) = 1293,05 + 0,000451122*xcoord(east) + 0,0544174*height - 0,0015277*Coastdistance





Homosedasticity is fullfilled, independence is not fulfilled.

Durbin-Watson statistic = 2,46851, this value should be near to 2.
Lag 1 residual autocorrelation = -0,238469, this value is not near to 0.
The transformation don't solve independence problem, only the homosedasticity problem.

3) Evaluation of the regression adjustment. Choice of the best multiple linear regression model

ModelResults

		Adjusted		Included
MSE	*R-Squared*	*R-Squared*	*Cp*	*Variables*
1086,0	0,0	0,0	201,275	
784,601	27,9991	27,7533	64,8825	A
998,922	8,33134	8,01848	162,095	B
1085,72	0,365727	0,0256785	201,467	C
955,773	12,291	11,9917	142,523	D
784,527	28,2515	27,7601	65,6346	AB
758,58	30,6245	30,1493	53,9056	AC
787,188	28,0082	27,5151	66,8372	AD
970,366	11,2557	10,6479	149,641	BC
842,451	22,9541	22,4264	91,8183	BD
812,707	25,6743	25,1653	78,373	CD
648,243	40,9183	40,3092	5,0259	ABC
769,27	29,8878	29,165	59,5471	ABD
721,349	34,2554	33,5776	37,9589	ACD
778,383	29,0572	28,3259	63,6522	BCD
645,966	41,3282	40,5189	5,0	ABCD

Models with Largest Adjusted R-Squared

		Adjusted		Included
MSE	*R-Squared*	*R-Squared*	*Cp*	*Variables*
645,966	41,3282	40,5189	5,0	ABCD
648,243	40,9183	40,3092	5,0259	ABC
721,349	34,2554	33,5776	37,9589	ACD
758,58	30,6245	30,1493	53,9056	AC
769,27	29,8878	29,165	59,5471	ABD
778,383	29,0572	28,3259	63,6522	BCD
784,527	28,2515	27,7601	65,6346	AB
784,601	27,9991	27,7533	64,8825	A
787,188	28,0082	27,5151	66,8372	AD
812,707	25,6743	25,1653	78,373	CD
842,451	22,9541	22,4264	91,8183	BD
955,773	12,291	11,9917	142,523	D
998,922	8,33134	8,01848	162,095	B
1085,72	0,365727	0,0256785	201,467	C
1086,0	0,0	0,0	201,275	

The best model includes all variables. When Cp statistic is equal 5,0 it is ideal model. It is better to choose model that include 4 variables.

ModelswithBestInformationCriteria

MSE	*Coefficients*	*AIC*	*HQC*	*SBIC*	*Included Variables*
648,243	4	6,50138	6,5214	6,55138	ABC
645,966	5	6,50465	6,52967	6,56714	ABCD
721,349	4	6,60824	6,62826	6,65823	ACD
758,58	3	6,65179	6,6668	6,68928	AC
769,27	4	6,67256	6,69258	6,72255	ABD
784,601	2	6,67873	6,68874	6,70373	A
778,383	4	6,68434	6,70436	6,73433	BCD
784,527	3	6,68542	6,70043	6,72291	AB
787,188	3	6,68881	6,70382	6,7263	AD
812,707	3	6,72071	6,73572	6,7582	CD
842,451	3	6,75665	6,77167	6,79415	BD
955,773	2	6,87608	6,88609	6,90108	D
998,922	2	6,92024	6,93024	6,94523	B
1086,0	1	6,99704	7,00204	7,00954	
1085,72	2	7,00356	7,01357	7,02856	C

The best model is that which include 3 variables. Two criterias indicate different best model.

		Standard	T	
Parameter	*Estimate*	*Error*	*Statistic*	*P-Value*
CONSTANT	2900,65	543,153	5,3404	0,0000
xcoord(east)	0,00222026	0,000285089	7,78795	0,0000
ycoord(north)	-0,000987919	0,000167087	-5,91261	0,0000
height	0,0638771	0,00849455	7,51978	0,0000
Coastdistance	0,000483279	0,000339538	1,42334	0,1557

Coastdistance can be eliminated from the model, because P-value >0,05

Analysis of Variance

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Model	131955,	4	32988,6	51,07	0,0000
Residual	187330,	290	645,966		
Total (Corr.)	319285,	294			

Explanatorymodel – P-value<0,05

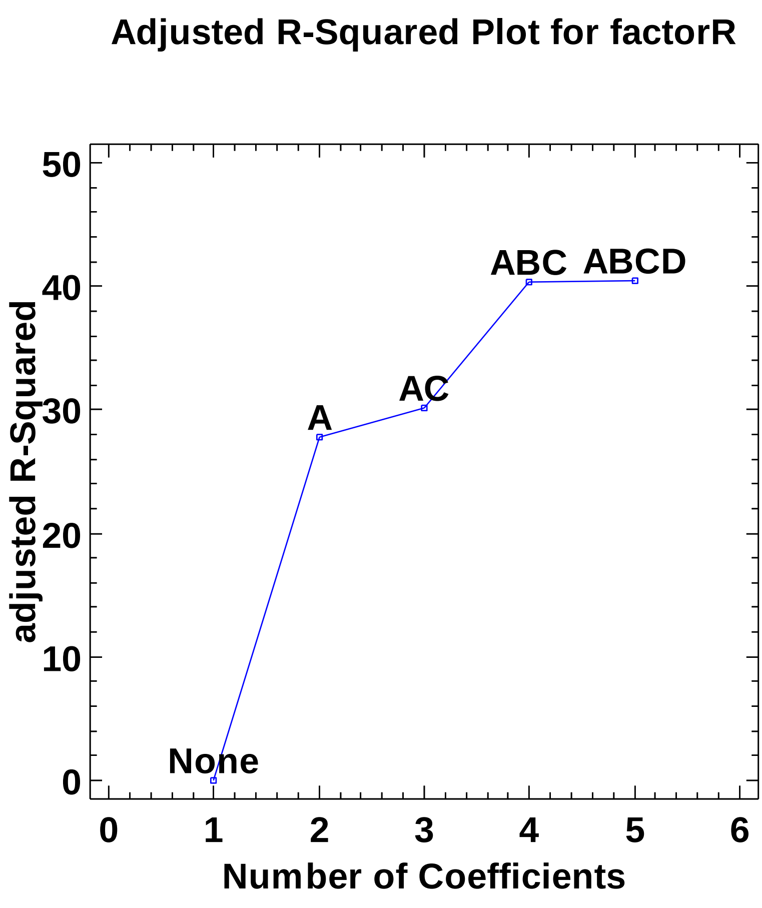
Model without CoastDistance

		Standard	T	
Parameter	*Estimate*	*Error*	*Statistic*	*P-Value*
CONSTANT	2388,72	407,72	5,85873	0,0000
xcoord(east)	0,00187994	0,000155531	12,0872	0,0000
ycoord(north)	-0,000814253	0,000114354	-7,12046	0,0000
height	0,0660843	0,00836653	7,89865	0,0000

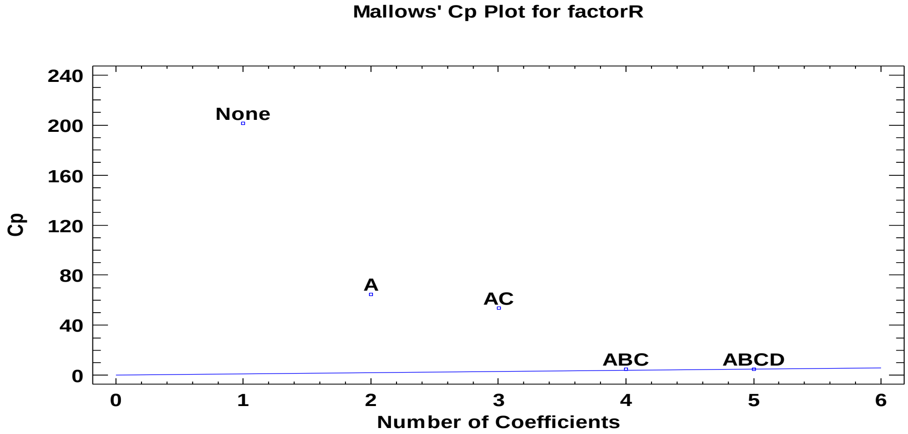
Analysis of Variance

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Model	130646,	3	43548,6	67,18	0,0000
Residual	188639,	291	648,243		
Total (Corr.)	319285,	294			

CONCLUSION: Model with 3 variables – xcoord, ycoord and heigh is considered as the best model.



The best model of each number of coefficients. Second best model include xcoord variable and height. The best model is that include all variable, but if we use only three first variable we have similar adjusted R-Squared like in the model which include all variables. We decided to choose both models.



4) Selection of variables for the regression model

Stepwise regression
Method: forward selection
P-to-enter: 0,05
P-to-remove: 0,05

Step 0:
0 variables in the model. 294 d.f. for error.
R-squared = 0,00%   Adjusted R-squared =   0,00%   MSE = 1086,0

Step 1:
Adding variable xcoord with P-to-enter =2,04881E-7
1 variables in the model. 293 d.f. for error.
R-squared = 28,00%   Adjusted R-squared = 27,75%   MSE = 784,601

Step 2:
Adding variable xcoord^2 with P-to-enter =0,00015195
2 variables in the model. 292 d.f. for error.
R-squared = 31,46%   Adjusted R-squared = 30,99%   MSE = 749,475

Step 3:
Adding variable height with P-to-enter =0,00072058
3 variables in the model. 291 d.f. for error.
R-squared = 34,10%   Adjusted R-squared = 33,42%   MSE = 723,021

Step 4:
Adding variable ycoord with P-to-enter =5,39169E-12
4 variables in the model. 290 d.f. for error.
R-squared = 44,08%   Adjusted R-squared = 43,31%   MSE = 615,669

Step 5:
Adding variable height^2 with P-to-enter =0,000758256
5 variables in the model. 289 d.f. for error.
R-squared = 46,24%   Adjusted R-squared = 45,31%   MSE = 593,985

Final modelselected.

The output shows the results of fitting a multiple linear regression model to describe the relationship between factorR and 7 independent variables. The equation of the fitted model is

factorR = -2288,02 + 0,0153076*xcoord - 0,000887486*ycoord + 0,131448*height - 9,03818E-9*xcoord^2 - 0,0000415707*height^2

Since the P-value in the ANOVA table is less than 0,05, there is a statistically significant relationship between the variables at the 95,0% confidence level.

5) Transformation of explanatory variables

Comparison of AlternativeModels

Model	*Correlation*	*R-Squared*
Doublereciprocal	0,5663	32,07%
Reciprocal-Y logarithmic-X	-0,5634	31,74%
Reciprocal-Y square root-X	-0,5618	31,56%
Reciprocal-Y	-0,5602	31,38%
Reciprocal-Y squared-X	-0,5568	31,00%
S-curve model	-0,5543	30,73%
Multiplicative	0,5514	30,40%
Logarithmic-Y square root-X	0,5498	30,23%
Exponential	0,5481	30,05%
Square root-Y reciprocal-X	-0,5457	29,78%
Logarithmic-Y squared-X	0,5447	29,67%
Square root-Y logarithmic-X	0,5427	29,45%
Doublesquareroot	0,5411	29,28%
Squareroot-Y	0,5395	29,10%
Square root-Y squared-X	0,5361	28,74%
Reciprocal-X	-0,5353	28,66%
Logarithmic-X	0,5323	28,34%
Squareroot-X	0,5308	28,17%
Linear	0,5291	28,00%
Squared-X	0,5257	27,64%
Squared-Y reciprocal-X	-0,5101	26,02%
Squared-Y logarithmic-X	0,5071	25,72%
Squared-Y square root-X	0,5056	25,56%
Squared-Y	0,5039	25,40%
Doublesquared	0,5005	25,05%
Logistic	<no fit>	
Log probit	<no fit>	

This table shows the results of fitting several curvilinear models to the data. Of the models fitted, the double reciprocal model yields the highest R-Squared value with 32,0699%. This is 4,07087% higher than the currently selected linear model. To change models, select the Analysis Options dialog box

Simple Regression - factorR vs. xcoord

Dependent variable: factorR
Independent variable: xcoord
Doublereciprocalmodel: **Y = 1/(a + b/X)**

Coefficients

	LeastSquares	*Standard*	***T***	
Parameter	*Estimate*	*Error*	***Statistic***	***P-Value***
Intercept	-0,0130943	0,00157841	**-8,2959**	**0,0000**
Slope	13571,8	1153,94	**11,7612**	**0,0000**

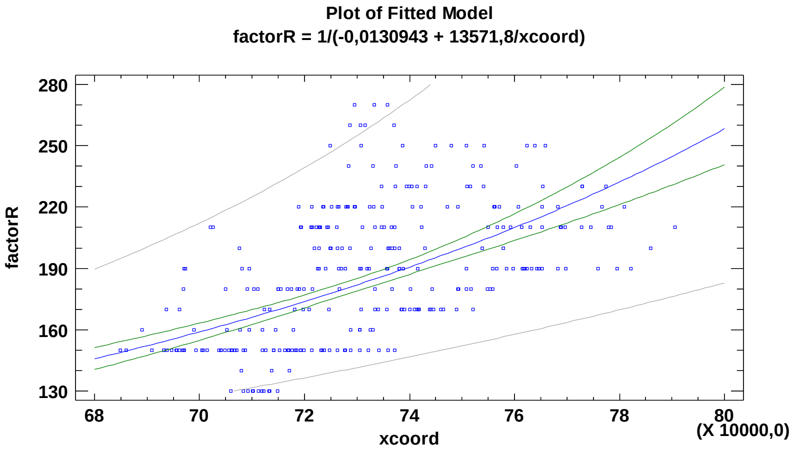
Analysis of Variance

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Model	0,0000882972	1	0,0000882972	138,33	0,0000
Residual	0,00018703	293	6,38327E-7		
Total (Corr.)	0,000275327	294			

CorrelationCoefficient = 0,566303
R-squared = 32,0699percent
R-squared (adjusted for d.f.) = 31,8381 percent
Standard Error of Est. = 0,000798954
Mean absolute error = 8939,65
Durbin-Watson statistic = 1,18435 (P=29,1038)
Lag 1 residual autocorrelation = 28,7357

The output shows the results of fitting a double reciprocal model to describe the relationship between factorR and xcoord. The equation of the fitted model is

factorR = 1/(-0,0130943 + 13571,8/xcoord)



Polynomial Regression - factorR versus xcoord

		Standard	T	
Parameter	*Estimate*	*Error*	*Statistic*	*P-Value*
CONSTANT	71641,6	41197,6	1,73898	0,0831
xcoord	-0,301323	0,168224	-1,7912	0,0743
xcoord^2	4,21767E-7	2,28842E-7	1,84305	0,0663
xcoord^3	-1,95992E-13	1,03707E-13	-1,88986	0,0598

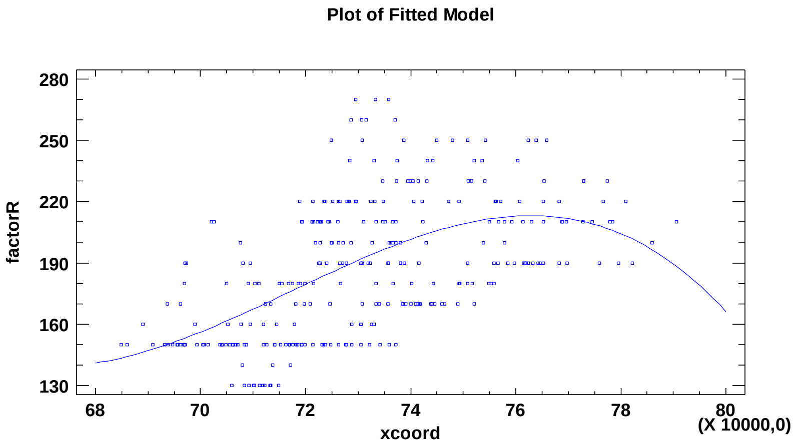
P-values are higher than 0,05, so this is not statistically significant polynomial. In polynomial regression options we change order to 3.

Analysis of Variance

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Model	103091,	3	34363,8	46,25	0,0000
Residual	216193,	291	742,932		
Total (Corr.)	319285,	294			

R-squared = 32,2882percent
R-squared (adjusted for d.f.) = 31,5902 percent
Standard Error of Est. = 27,2568
Mean absolute error = 22,1796
Durbin-Watson statistic = 1,10534 (P=0,0000)
Lag 1 residual autocorrelation = 0,443263

EQUATION: factorR = 71641,6-0,301323*xcoord + 4,21767E-7*xcoord^2-1,95992E-13*xcoord^3



Unrealistic for our data.

Polynomial Regression - factorR versus xcoord

		Standard	T	
Parameter	*Estimate*	*Error*	*Statistic*	*P-Value*
CONSTANT	-6164,74	1503,47	-4,10035	0,0001
xcoord	0,0165048	0,00409199	4,03344	0,0001
xcoord^2	-1,06806E-8	2,7827E-9	-3,83822	0,0002

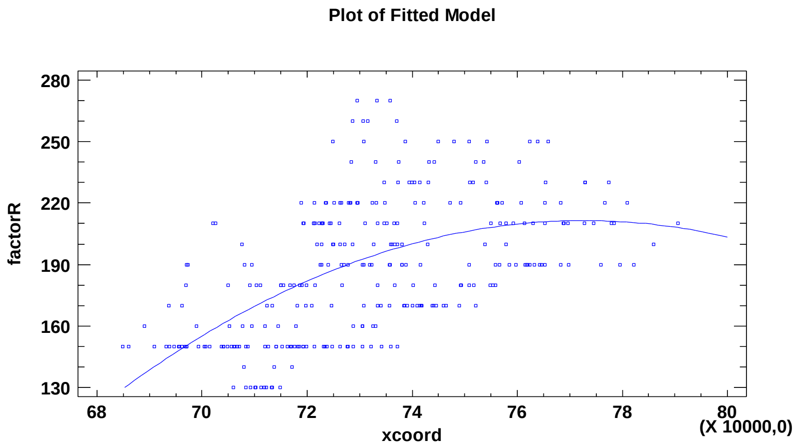
P-values are lower than 0,05, so this is statistically significant polynomial. In polynomial regression options we have to change order to 2.

Analysis of Variance

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Model	100438,	2	50219,0	67,01	0,0000
Residual	218847,	292	749,475		
Total (Corr.)	319285,	294			

R-squared = 31,4572percent
R-squared (adjusted for d.f.) = 30,9877 percent
Standard Error of Est. = 27,3765
Mean absolute error = 22,3218
Durbin-Watson statistic = 1,13776 (P=0,0000)
Lag 1 residual autocorrelation = 0,427576

EQUATION: factorR = -6164,74 + 0,0165048*xcoord-1,06806E-8*xcoord^2



Stepwise regression
Method: forward selection
P-to-enter: 0,05
P-to-remove: 0,05

Step 0:
0 variables in the model. 294 d.f. for error.
R-squared = 0,00%   Adjusted R-squared =   0,00%   MSE = 1086,0

Step 1:
Adding variable xcoord with P-to-enter =2,04881E-7
1 variables in the model. 293 d.f. for error.
R-squared = 28,00%   Adjusted R-squared = 27,75%   MSE = 784,601

Step 2:
Adding variable xcoord^2 with P-to-enter =0,00015195
2 variables in the model. 292 d.f. for error.
R-squared = 31,46%   Adjusted R-squared = 30,99%   MSE = 749,475

Step 3:
Adding variable height with P-to-enter =0,00072058
3 variables in the model. 291 d.f. for error.
R-squared = 34,10%   Adjusted R-squared = 33,42%   MSE = 723,021

Step 4:
Adding variable ycoord with P-to-enter =5,39169E-12
4 variables in the model. 290 d.f. for error.
R-squared = 44,08%   Adjusted R-squared = 43,31%   MSE = 615,669

Step 5:
Adding variable height^2 with P-to-enter =0,000758256
5 variables in the model. 289 d.f. for error.
R-squared = 46,24%   Adjusted R-squared = 45,31%   MSE = 593,985

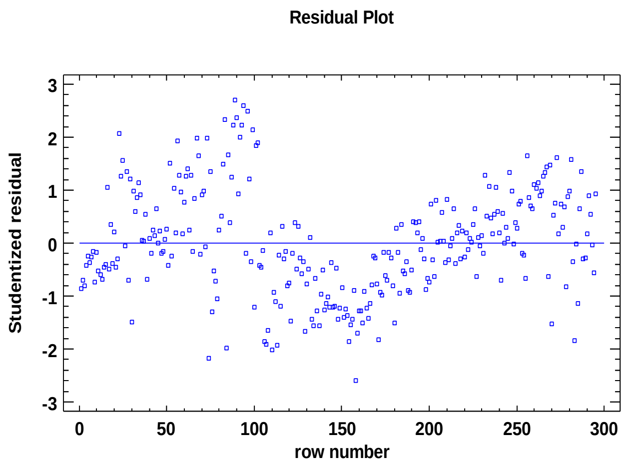
Final model selected.

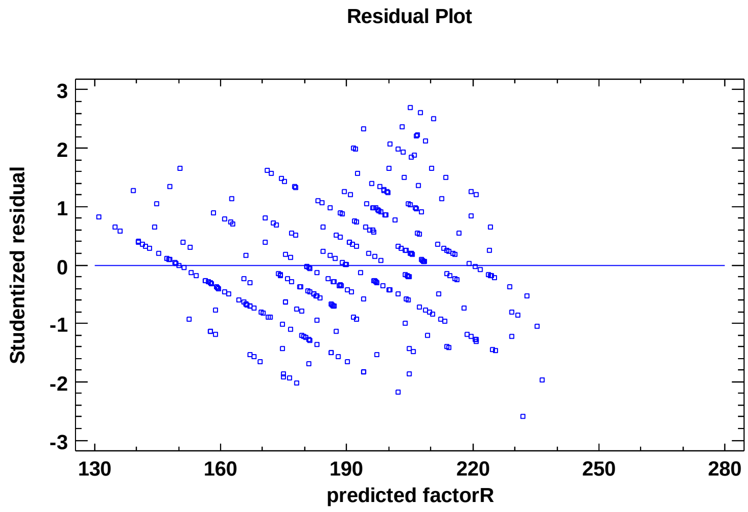
The StatAdvisor
The output shows the results of fitting a multiple linear regression model to describe the relationship between factorR and 7 independent variables. The equation of the fitted model is

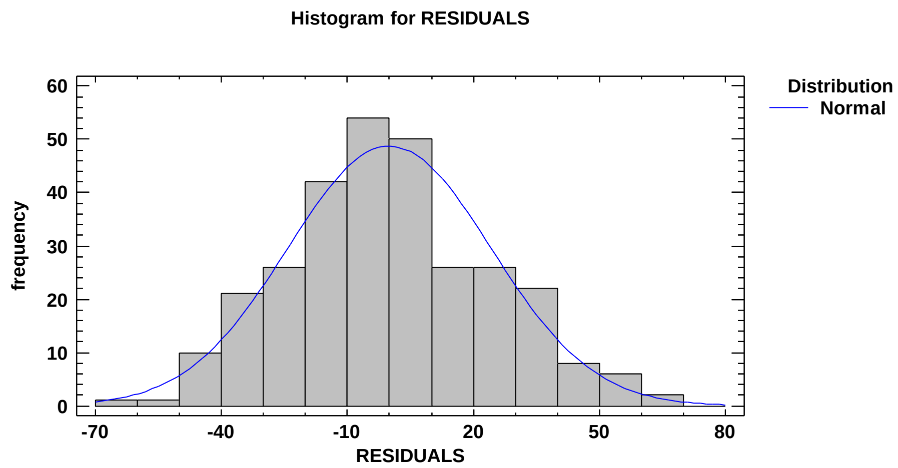
factorR = -2288,02 + 0,0153076*xcoord - 0,000887486*ycoord + 0,131448*height - 9,03818E-9*xcoord^2 - 0,0000415707*height^2

Since the P-value in the ANOVA table is less than 0,05, there is a statistically significant relationship between the variables at the 95,0% confidence level.

R-squared (adjusted for d.f.) = 45,3054 percent
Standard Error of Est. = 24,3718







Comparing the obtained results to these from previous parts of exercise, this model improved some of the indicators, for example data better follow normal distribution. But the rest of the obtained results are very similar.