

X1: Vzdálenost od rostliny v metrech.

Y2: Počet semen na povrchu půdy.

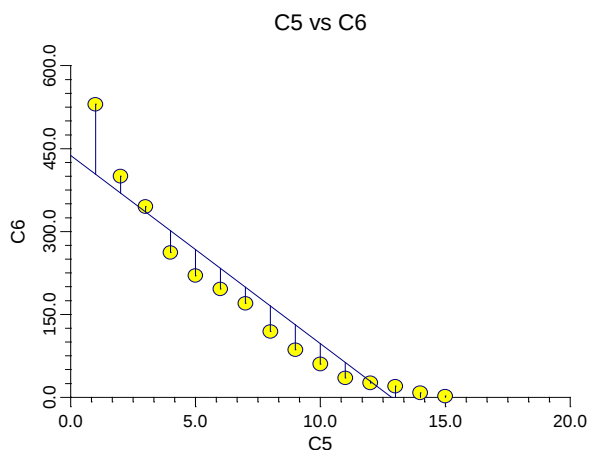
Y2	X1
530	1
400	2
345	3
262	4
220	5
196	6
170	7
119	8
86	9
60	10
35	11
26	12
20	13
8	14
2	15

Příklad 14: Kvadratická závislost v datech, linearizace dat.

Použité proměnné: X1, Y2

Studujeme závislost disperze semen od mateřské rostliny a měříme počet semen spadných na povrch půdy v závislosti na vzdálenosti od mateřské rostliny. Získaná data vypadají graficky následovně. Závislost se zdá být lineární ve střední části, na obou koncích jsou pak odchylky od přímky větší.

Scatter Plot Section



Použijeme-li lineární regresi, získáme následující rovnici: $y = 437.5 - 34x$. Počet semen tedy velmi silně klesá se vzdáleností od rostliny. Regresní koeficient b je průkazně odlišný od nuly, koeficient determinace vysvětluje téměř 90% variability dat, ANOVA modelu je (pochopitelně, viz předchozí příklad) taktéž velmi průkazná. Zdálo by se tudíž, že vše je v naprostém pořádku.

Multiple Regression Report

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	437.4952	29.33465	14.9139	0.000000	Reject Ho	1.000000
C2	-34.02857	3.226381	-10.5470	0.000000	Reject Ho	1.000000
R-Squared	0.895363					

Analysis of Variance Section

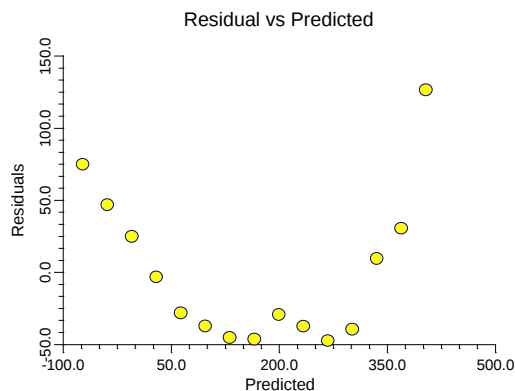
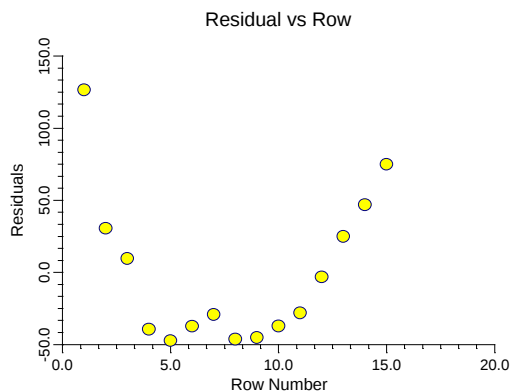
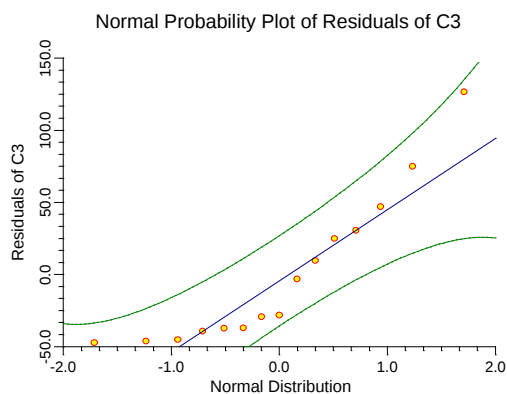
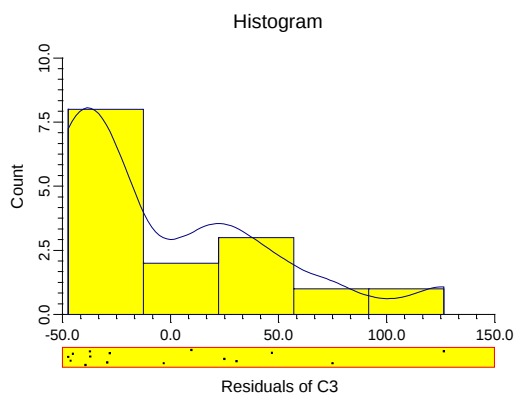
Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	409696.1	409696.1			
Model	1	324224.2	324224.2	111.2388	0.000000	1.000000
Error	13	37890.7	2914.67			
Total(Adjusted)	14	362114.9	25865.35			
Root Mean Square Error		53.98768	R-Squared	0.8954		
Mean of Dependent		165.2667	Adj R-Squared	0.8873		
Coefficient of Variation		0.3266701	Press Value	58320.09		
Sum Press Residuals		745.9579	Press R-Squared	0.8389		

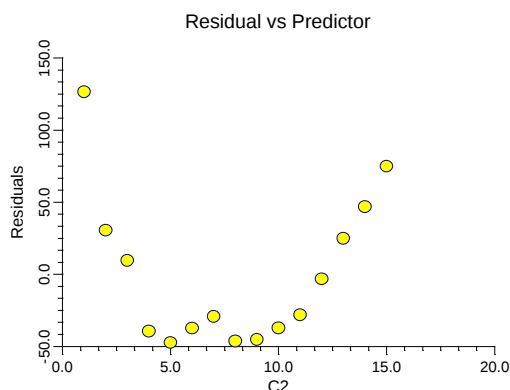
Normality Tests Section

Assumption	Value	Probability	Decision(5%)
Skewness	2.0604	0.039361	Rejected
Kurtosis	1.0004	0.317103	Accepted
Omnibus	5.2461	0.072582	Accepted

Problémem ale je skutečnost, že náš předpoklad lineární závislosti nemusí být vůbec opodstatněný. Prohlídkou reziduálů zjistíme, že jejich rozdělení není vůbec náhodné, naopak jejich uspořádání vykazuje jednoznačný trend (reziduály tvoří jakousi parabolu). Předpoklad lineární závislosti v datech a tudíž použití lineární regrese v tomto příkladu je totiž chybné a nic na tom nezmění ani zjištěná průkaznost výsledků. Zde je velmi názorné vidět, že je třeba **vždy** zkontrolovat rozdělení reziduálů a tedy ověřit oprávněnost použitého modelu!

Plots Section





V uvedeném příkladě se dá očekávat, že s rostoucí vzdáleností od zdroje bude počet semen ubývat spíše podle kvadratické závislosti. Jsou dvě možnosti jak toto ověřit, buďto použít kvadratický model nebo příslušnou transformací zlinearizovat data a poté s takto upravenými daty znovu provést lineární regresi. Zde použijeme druhou variantu.

Transformace dat v NCSS se zadává v dialogové okně spreadsheetu (červená kostka) v možnostech **Data/Transformations**. Nejdříve je třeba zadat sloupeček s cílovou proměnnou (*Result Variable*), pak požadovanou transformační funkci (*Available Functions*) a nakonec transformovanou proměnnou (*Argument Variable*). Po odklepnutí OK je třeba vlastní transformaci provést v **Edit/Transform Current**.

Následující lineární regrese je provedena s odmocninou transformací proměnné Y2. Regresní model má nyní podobu $y = 22.9 - 1.47x$, regresní koeficient je opět průkazně odlišný od nuly a tedy i ANOVA je průkazná. Především se ale markantně zvýšil koeficient determinace, $R^2 = 99\%$. To znamená, že tento model vysvětluje mnohem více variability dat než model předchozí.

Multiple Regression Report

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	22.93019	0.3717511	61.6816	0.000000	Reject Ho	1.000000
C2	-1.472013	4.088717E-02	-36.0018	0.000000	Reject Ho	1.000000
R-Squared	0.990070					

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	1866.205	1866.205			
Model	1	606.7102	606.7102	1296.1316	0.000000	1.000000
Error	13	6.08521	0.4680931			
Total(Adjusted)	14	612.7954	43.7711			

Root Mean Square Error	0.6841733	R-Squared	0.9901
Mean of Dependent	11.15409	Adj R-Squared	0.9893
Coefficient of Variation	6.133835E-02	Press Value	9.068867
Sum Press Residuals	8.982198	Press R-Squared	0.9852

Normality Tests Section

Assumption	Value	Probability	Decision(5%)
Skewness	1.3883	0.165049	Accepted
Kurtosis	0.8445	0.398412	Accepted
Omnibus	2.6405	0.267074	Accepted

Následná inspekce rozdělení reziduálů potvrzuje skutečnost, že nyní je již lineární model použit oprávněně.

Nežádoucí trend v rozdělení reziduálů byl vhodnou (kvadratickou) transformací dat odstraněn a jejich rozdělení je v “požadovaném” pásu okolo nulové hodnoty.

Plots Section

