X1: Počet dní pokusu.
Y1: Výška rostliny (v cm) od vyklíčení.

Y1	X1
1.5	1
3	2
4.5	3
5	4
6	5
8	6
9	7
11	8
13	9
14	10
15	11
16	12
18.5	13
20	14
23	15

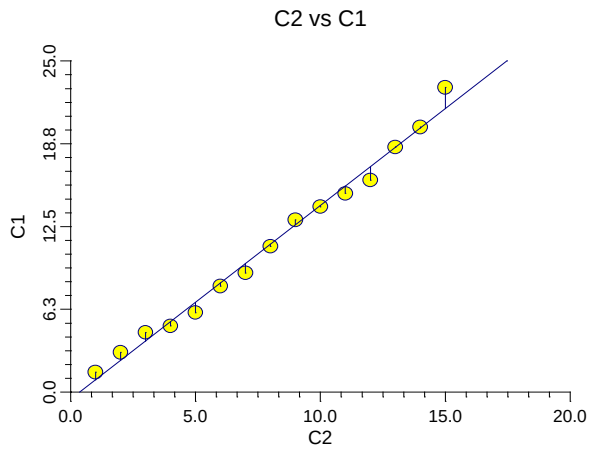
Příklad 13: Jednoduchá lineární regrese.

Použité proměnné: X1, Y1

Vztah dvou nebo více kontinuálních proměnných je řešen regresní analýzou. V regresní analýze obecně platí, že jedna (či více) proměnná je nezávislá (*x*), pomocí které se vysvětluje proměnná závislá (*y*). Obecným předpokladem je skutečnost, že proměnnou *x* měříme s nulovou chybou, nebo alespoň její chyba je oproti chybě proměnné *y* velmi malá. Jsou-li proměnné pouze dvě a předpokládáme-li lineární závislost, jedná se o jednoduchou lineární regresi, jako nejjednodušší typ regrese. Vyjádřením vztahu obou proměnných je potom regresní rovnice tvaru $y = a + b.x$ Cílem regresní analýzy je nalezení parametrů *a* (průsečík s osou y) a *b* (regresní koeficient, směrnice regresní přímky).

V našem příkladu studujeme závislost růstu rostliny na čase a měříme její výšku s počtem dní pokusu. Protože je k dispozici pouze jedna nezávislá proměnná (čas), jedná se o jednoduchou lineární regresi. Graficky vztah obou proměnných zobrazíme v oddíle **Graphics/Scatter Plots**. Pokud v části *Trend Line/Type of Line* zvolíme možnost *Least Square* a v oddíle *Residuals/Show Residuals* zvolíme *Yes*, získáme následující graf, který kromě zobrazení jednotlivých bodů jimi prokládá přímku (podle metody nejmenších čtverců) a zároveň zobrazuje *reziduály* (tj. rozdíl mezi naměřenou a predikovanou hodnotou).

Scatter Plot Section



Volbou **Analysis/Regression/Correlation/Multiple Regression** voláme regresní analýzu, kde zadáme příslušné vysvětlující a vysvětlované proměnné (a protože vysvětlující proměnná je pouze jedna, jedná se o

jednoduchou regresi). Výsledkem lineární regrese proměnných Y1 a X1 je následující regresní rovnice **y = –0.53 + 1.46x**. Řádek uvedený jako *Intercept* odpovídá parametru *a*, a řádek uvedený jako *C2* odpovídá parametru *b*. Stěžejní otázkou regresní analýzy je otázka, jestli je regresní koeficient (*Regression Coefficient*, parametr *b*) průkazně odlišný o nuly, jinými slovy existuje-li statistická závislost mezi proměnnými *x* a *y*. V tomto případě je na základě hodnoty koeficientu *b*, jeho střední chyby (*Standard Error*) a následně příslušného T-testu (*T-value*) rozhodnuto zamítnout nulovou hypotézu, že regresní koeficient není odlišný od nuly.

Parament *R-Squared*, tzv. koeficient determinace R^2, je měřítkem vysvětlující síly regresního modelu (regresní rovnice) a udává podíl variability vysvětlené regresním modelem vůči variabilitě celkové. V našem případě můžeme konstatovat, že téměř 99% celkové variability dat bylo vysvětleno regresním modelem (charakterizovaným regresní rovnicí).

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	-0.5333334	0.3824225	-1.3946	0.186506	Accept Ho	0.252901
C2	1.4625	4.206086E-02	34.7710	0.000000	Reject Ho	1.000000
R-Squared	0.989362					

Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	-0.5333334	0.3824225	-1.359507	0.2928402	0.0000
C2	1.4625	4.206086E-02	1.371633	1.553367	0.9947
T-Critical	2.160369				

Výpočet regresní rovnice doprovází také analýza variance příslušného regresního modelu. Analýza variance regresního modelu obecně je testem, který zjišťuje, zda-li regresní model vysvětluje signifikantní část variability dat. Jedná se o podíl variance vysvětlené regresním modelem (řádek *Model*, sloupeček *Mean Square*) vůči varianci zbytkové (řádek *Error*, sloupeček *Mean Square*). Získaná hodnota F-statistiky při daném počtu stupňů odpovídá míře průkaznosti regresního modelu. Na rozdíl od koeficientu determinace R^2 (ten je dán podílem sumy čtverců v důsledku modelu (řádek *Model*, sloupeček *Sum of Squares*) a celkové sumy čtverců (řádek *Total*, sloupeček *Sum of Squares*)) tak ANOVA dává rigorózní test průkaznosti modelu.

V případě jednoduché regrese, analýza variance regresního modelu v podstatě testuje nulovou hypotézu, je-li regresní koeficient *b* průkazně odlišný od nuly (a je tedy analogická příslušnému T-testu). U mnohorozměrných regresních modelů, kde je více parciálních regresních koeficientů, pak analýza variance testuje průkaznost celého modelu.

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	1870.417	1870.417			
Model	1	598.8937	598.8937	1209.0252	0.000000	1.000000
Error	13	6.439583	0.4953526			
Total(Adjusted)	14	605.3333	43.23809			

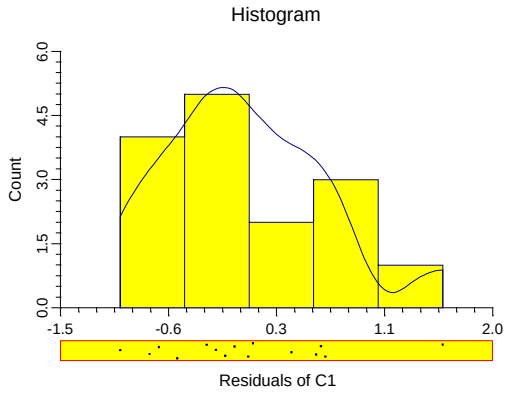
Root Mean Square Error	0.7038129	R-Squared 0.9894	
Mean of Dependent	11.16667	Adj R-Squared	0.9885
Coefficient of Variation	6.302802E-02	Press Value	9.604154
Sum |Press Residuals|	9.174129	Press R-Squared	0.9841

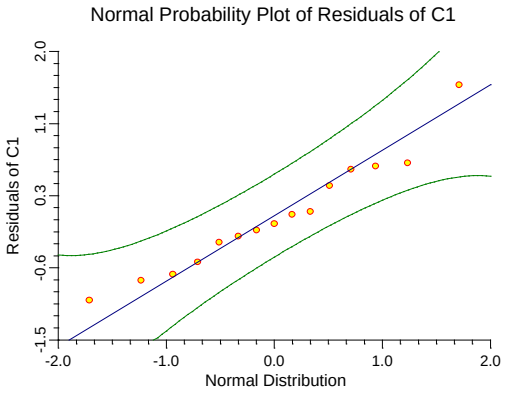
Normality Tests Section

Assumption	Value	Probability	Decision(5%)
Skewness	1.2639	0.206277	Accepted
Kurtosis	0.8566	0.391659	Accepted
Omnibus	2.3312	0.311743	Accepted

Jedním z předpokladů regresní analýzy, podobně jako u ANOVy, je normalita rozdělení reziduálů. Použití lineárního modelu je oprávněné tehdy, je-li studovaná závislost v datech (přibližně) lineární. Pokud tomu tak opravdu je, rozdělení rozdělení reziduálů nevykazuje žádný trend a blíží se normalitě. Graficky je rozdělení reziduálů zachyceno na následujících grafech a jejich vizuální inspekce je důležitá pro ověření správnosti použitého regresního modelu.

Plots Section





Ideální rozdělení reziduálů vůči predikované nebo vysvětlující proměnné je v rovnoměrném pásu okolo nulové hodnoty (jen pro úplnost, suma hodnot reziduálů je rovna nule), což je i tento případ. Po prohlídce rozdělení reziduálů můžeme s klidným svědomím konstatovat, že použitý lineární model je v tomto případě adekvátní.