Example: Ice Cream Sales

The local ice cream shop keeps track of how much ice cream they sell versus the temperature of that day for t
Formulate a null hypothesis and verify it by Pearsons and Spearman coefficients

Temperature (°C)	Ice Cream Sales ($)
14.2	215
16.4	325
11.9	185
15.2	332
18.5	406
22.1	522
19.4	412
25.1	614
23.4	544
18.1	421
22.6	445
17.2	408

$r =$

:he last 12 days:

$$r_S = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n}$$

$n \backslash \alpha$	0.2	0.1	0.05
4	1.000	1.000	—
5	0.800	0.900	1.000
6	0.657	0.829	0.886
7	0.571	0.714	0.786
8	0.524	0.643	0.738
9	0.483	0.600	0.700
10	0.455	0.564	0.648
11	0.427	0.536	0.618
12	0.406	0.503	0.587
13	0.385	0.484	0.560
14	0.367	0.464	0.538
15	0.354	0.446	0.521
16	0.341	0.429	0.503
17	0.328	0.414	0.488

$$= \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum (X_i - \overline{X})^2 \sum (Y_i - \overline{Y})^2}}$$

Pearson r crit.	One-Tailed Test		
	.05	.025	.01
	Two-Tailed Test		
df	.10	.05	.02
1	.988	.997	.9995
2	.900	.950	.980
3	.805	.878	.934
4	.729	.811	.882
5	.669	.754	.833
6	.622	.707	.789
7	.582	.666	.750
8	.549	.632	.716
9	.521	.602	.685
10	.497	.576	.658

0.02	0.01	n \ α	0.2	0.1	0.05	0.02	0.01
—	—	18	0.317	0.401	0.472	0.550	0.600
1.000	—	19	0.309	0.391	0.460	0.535	0.584
0.943	1.000	20	0.299	0.380	0.447	0.522	0.570
0.893	0.929	21	0.292	0.370	0.436	0.509	0.556
0.833	0.881	22	0.284	0.361	0.425	0.497	0.544
0.783	0.833	23	0.278	0.353	0.416	0.486	0.532
0.745	0.794	24	0.271	0.344	0.407	0.476	0.521
0.709	0.755	25	0.265	0.337	0.398	0.466	0.511
0.678	0.727	26	0.259	0.331	0.390	0.457	0.501
0.648	0.703	27	0.255	0.324	0.383	0.449	0.492
0.626	0.679	28	0.250	0.318	0.375	0.441	0.483
0.604	0.654	29	0.245	0.312	0.368	0.433	0.475
0.582	0.635	30	0.240	0.306	0.362	0.425	0.467
0.566	0.618	rho critical values for 2-tailed test					

age (yrs)	price/1000 Kč
3	167
4	165
5	139
6	149
7	119
7	129
8	89
8	115
9	76
9	89

Here is a pricelist of used 10 cars Skoda Felicia Combi

1. pressume normal distribution of the data
2. construct a simple regression model how the price depends on th
3. evaluate quality of the model
4. estimate a price of a ten-year-old Felicia Combi

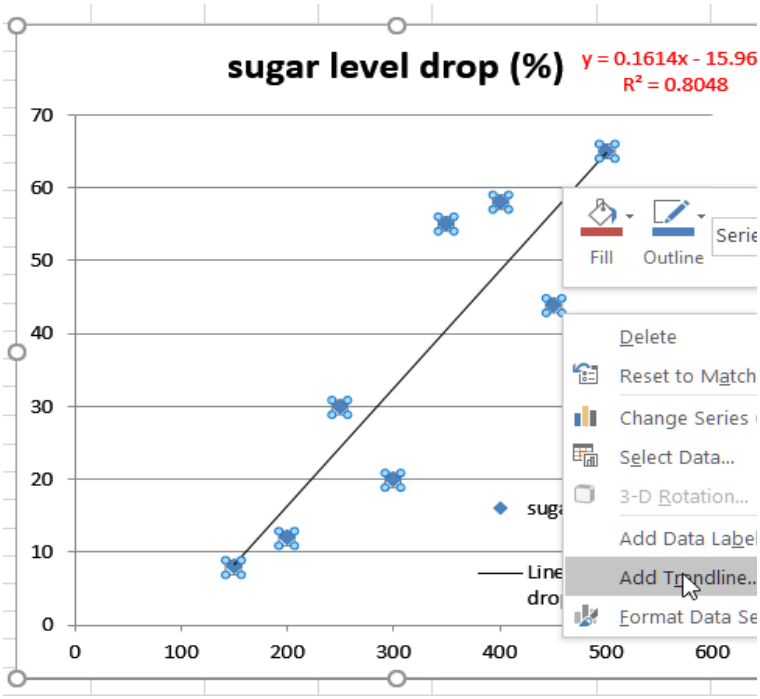
e age

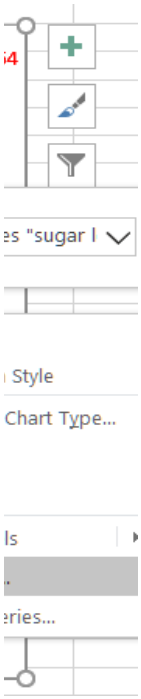
A new kind of insulin was developed. Its effect was tested as a drop of sugar level in blood 2 hours after the injection application.
8 Randomly selected patients were dozed with different insulin amounts.
Results are in the table: **Prove a strong correlation and plot a graph of regression residua**

insuline amount (ug)	150	200	250	300	350	400
sugar level drop (%)	8	12	30	20	55	58

als!

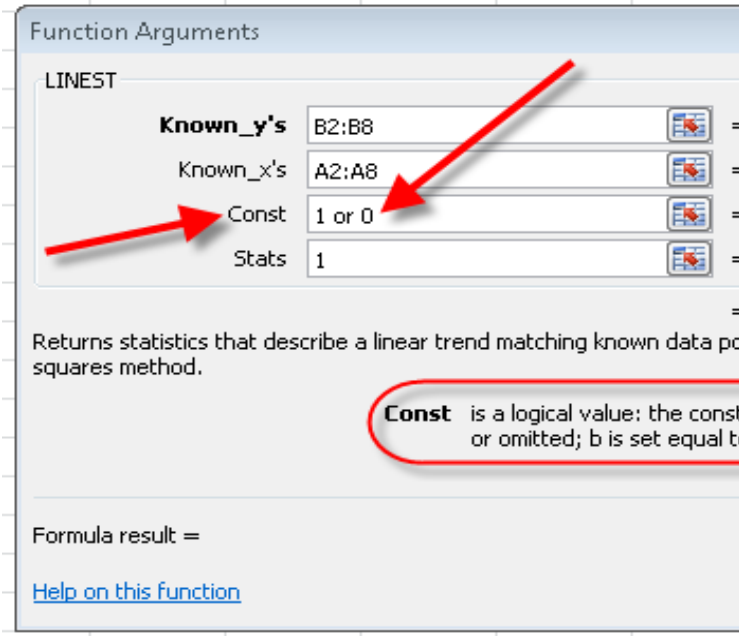
450	500
44	65

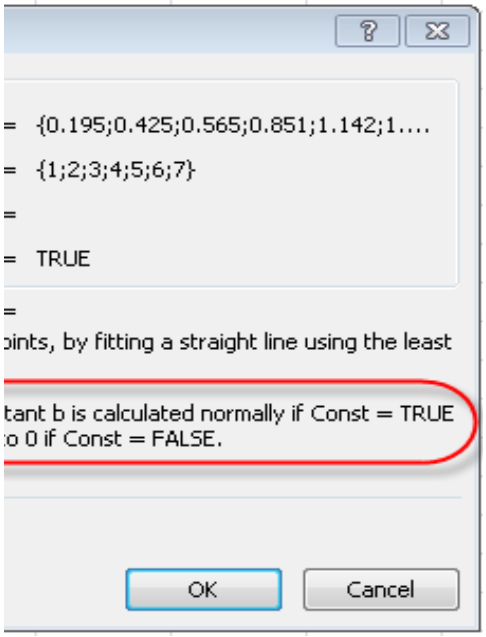


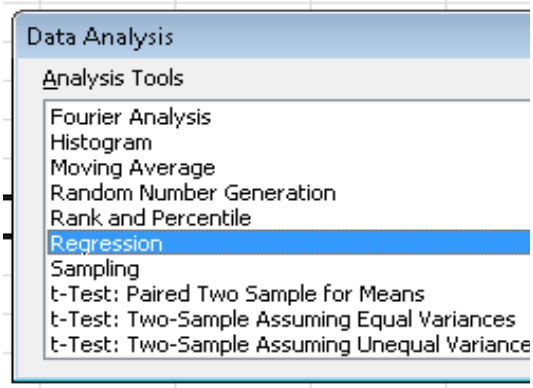

es "sugar l
Style
Chart Type...
ls
eries...

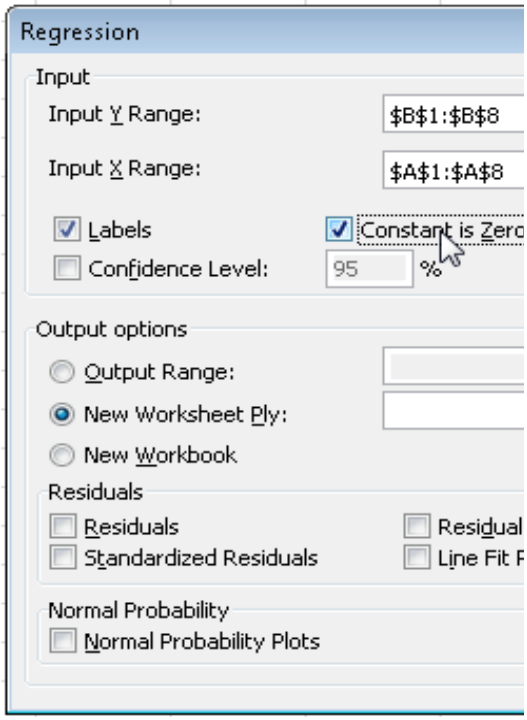
concentration	signal
1	0.195
2	0.425
3	0.565
4	0.851
5	1.142
6	1.198
7	1.530

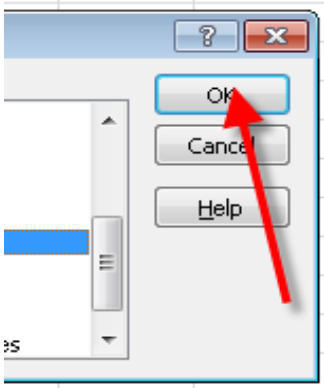
HOW TO FORCE Const a=0

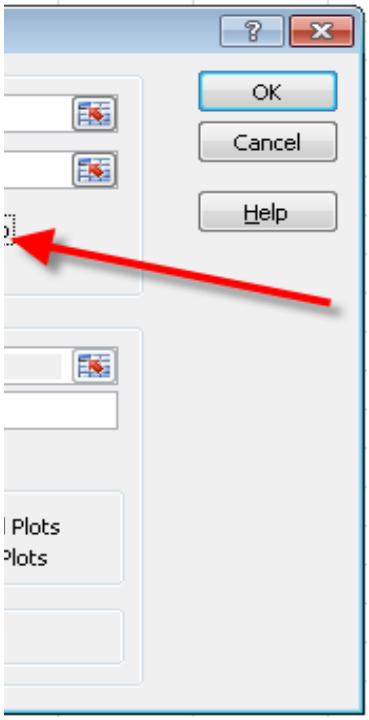



= {0.195;0.425;0.565;0.851;1.142;1....
= {1;2;3;4;5;6;7}
=
= TRUE
=
oints, by fitting a straight line using the least
tant b is calculated normally if Const = TRUE
o 0 if Const = FALSE.
OK
Cancel


Data Analysis
Analysis Tools
Fourier Analysis
Histogram
Moving Average
Random Number Generation
Rank and Percentile
Regression
Sampling
t-Test: Paired Two Sample for Means
t-Test: Two-Sample Assuming Equal Variances
t-Test: Two-Sample Assuming Unequal Variance


Regression
Input
Input Y Range:
$B$1:$B$8
Input X Range:
$A$1:$A$8
Labels
Constant is Zero
Confidence Level:
95
%
Output options
Output Range:
New Worksheet Ply:
New Workbook
Residuals
Residuals
Standardized Residuals
Residual
Line Fit P
Normal Probability
Normal Probability Plots


OK
Cancel
Help

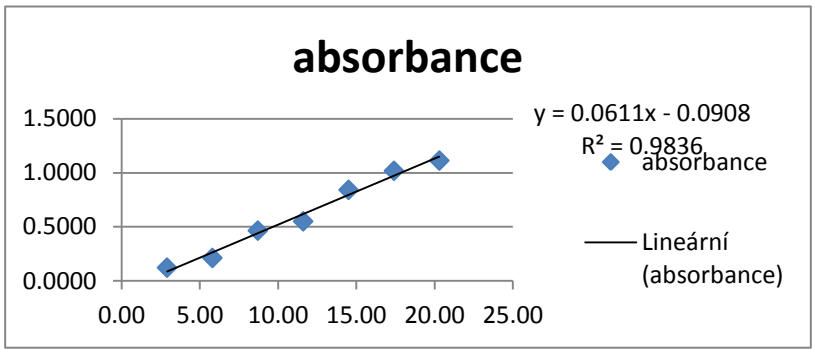

OK
Cancel
Help
Plots
Plots

mmol/L

concentration	absorbance
2.90	0.1225
5.80	0.2125
8.70	0.4650
11.60	0.5510
14.50	0.8420
17.40	1.0184
20.30	1.1130

1/
2/
3/
b (
a (
Pe
co
4/

= linest() / =linregrese()



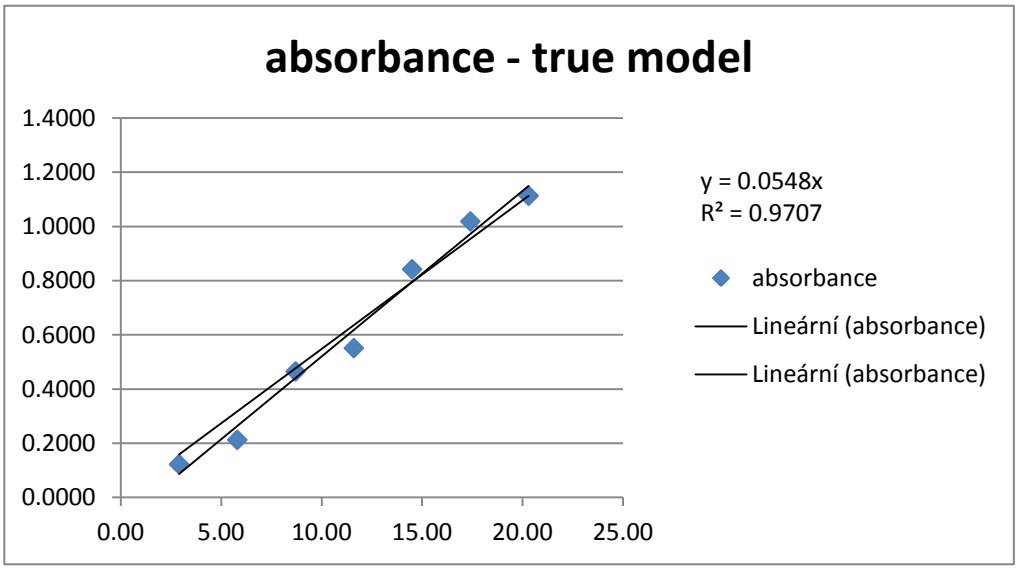
5x2 matice!!!

0.061087438
0.003525542
0.98361883
300.2285106
0.87873486

>>
>>
>>

>>

H0: "a" is a 0
verification:
>> perform the standard deviation (sample) and simple t-test and compare it with Tcrit (alpha; n-2
>> n-2 is degrees of freedom
>> t = a/Sa
>> if we will not overstep critical value then we accept null hypothesis and model should simplify t



>> Evaluation how good the model is

VÝSLEDEK

Regresní statistika	
Násobné R	0.996320178
Hodnota spolehlivosti R	0.992653897
Nastavená hodnota spolehlivosti R	0.825987231
Chyba stř. hodnoty	0.066065408
Pozorování	7

ANOVA

	Rozdíl	*SS*
Regrese	1	3.538672232
Rezidua	6	0.026187829
Celkem	7	3.56486006

	Koeficienty	*Chyba stř. hodnoty*
Hranice	0	#N/A
concentration	0.054822414	0.001925361

REZIDUA

Pozorování	*Očekávané absorbance*	*Rezidua*
1	0.158985	-0.036485
2	0.31797	-0.10547
3	0.476955	-0.011955
4	0.63594	-0.08494
5	0.794925	0.047075
6	0.95391	0.06449
7	1.112895	0.000105

perform a linear regression of the data in the form Y=b*X+a, incl. a graph
test a null hypothesis H0: a=0
based on your conclusion calculate the following 4 numbers:
(2 decimal digits)
(2 dec. digits)
earson´s correlation coeficient r = sqrt(R^2) (4 decimal digits)
ncentration of unknown, if absorbance of it was 0.502
submit in an Excel sheet incl. a correct calibration graph

		t	5x2 matice!!!	
-0.090842857		1.98678943	**b**	**a**
0.045723445			sb	sa
0.05410071		tcrit	R2	syx
5		2.570581836	F	dof
0.014634434			ssreg	ssres

1,987 < 2,571

H0 is accepted

"a" is a random value if we perform repeat the analysis with another 8 patients
maybe the intercept would be positive or maybe less negative
whatever so this has no meaning so we should immediately after establishing the hypothesis o
test if this is significant and if not we should repeat the analysis actually with trendline set in zer
we get statistically better quation decribing the model

2)

hat would be 0 and the function would be y is equal to P times

Do not use that kind of R2 (R2=0,9707), when we make the regression line, in case we simplify

or are right values

5x2 matice!!!

	0.054822414	0
	0.001925361	#N/A
R2	0.992653897	0.066065408
R = 0.996320178	810.7596011	6
	3.538672232	0.026187829

R - quality of the model

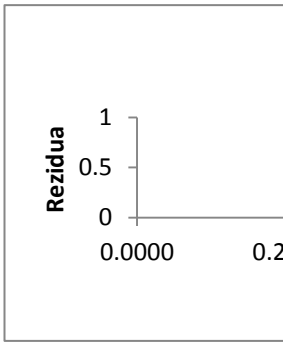
Absorbance = 0.502
Concentration = 9.1568

Pearsons corralation coeficient

MS	*F*	*Významnost F*
3.538672232	810.7596011	1.00081E-06
0.004364638		

t Stat	*Hodnota P*	*Dolní 95%*	*Horní 95%*
#N/A	#N/A	#N/A	#N/A
28.47384065	1.24228E-07	0.050111226	0.059533602

f this regression line we should
˙o



y the model (a=0)

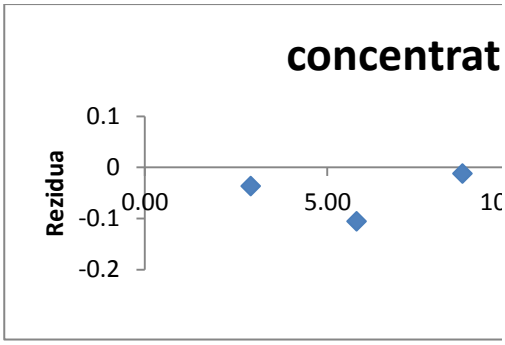
5x2 matice!!!

b	a
sb	sa
R2	syx
F	dof
ssreg	ssres

Dolní 95,0%	*Horní 95,0%*
#N/A	#N/A
0.050111226	0.059533602

VÝSLEDEK

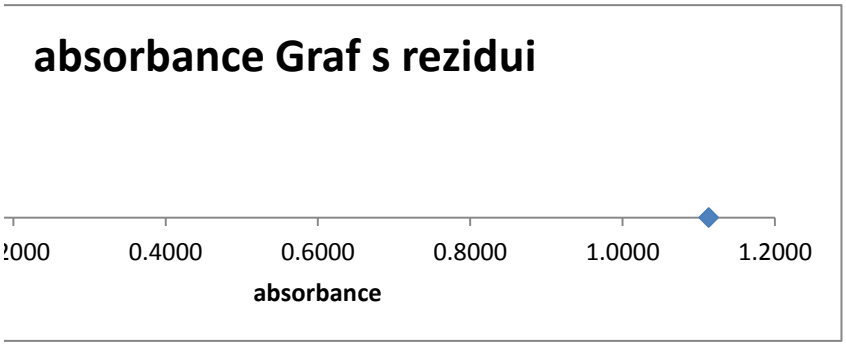
Regresní statistika	
Násobné R	0.996494119
Hodnota spolehlivosti R	0.99300053
Nastavená hodnota spolehlivosti R	0.79300053
Chyba stř. hodnoty	0.070494108
Pozorování	6

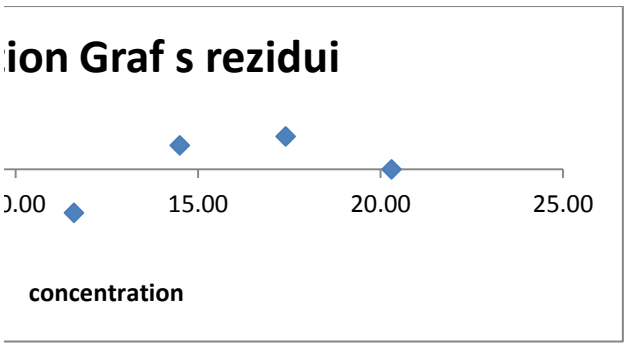


ANOVA

	Rozdíl	*SS*	*MS*
Regrese	1	3.525006713	3.525006713
Rezidua	5	0.024847097	0.004969419
Celkem	6	3.54985381	

		Koeficienty	Chyba stř. hodnoty	t Stat
Hranice		0	#N/A	#N/A
	2.9	0.054912925	0.002061804	26.63343314





F	*Významnost F*
709.3397606	1.18133E-05

Hodnota P	Dolní 95%	Horní 95%	Dolní 95,0%
#N/A	#N/A	#N/A	#N/A
1.39518E-06	0.049612888	0.060212961	0.049612888

Horní 95,0%
#N/A
0.060212961