6. Lineární regresní modely

- 6.1 Jednoduchá regrese a validace
- 6.2 Testy hypotéz v lineární regresi
- 6.3 Kritika dat v regresním tripletu
- 6.4 Multikolinearita a polynomy
- 6.5 Kritika modelu v regresním tripletu
- 6.6 Kritika metody v regresním tripletu
- 6.7 Lineární a nelineární kalibrace
- 7. Korelační modely

Cíl regresní analýzy

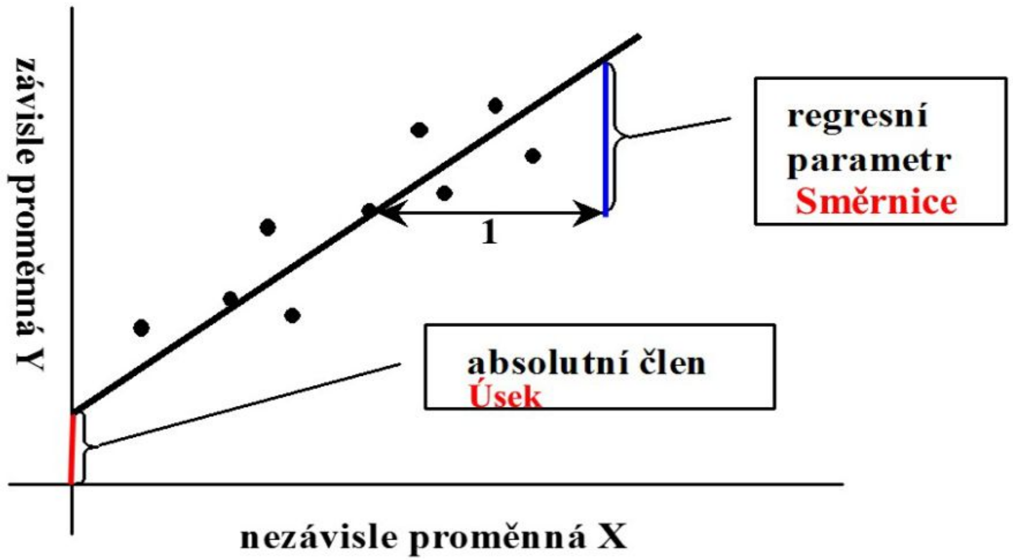
Cílem regresní analýzy je nalezení vhodného modelu studované závislosti tak, že se **snažíme nahradit**

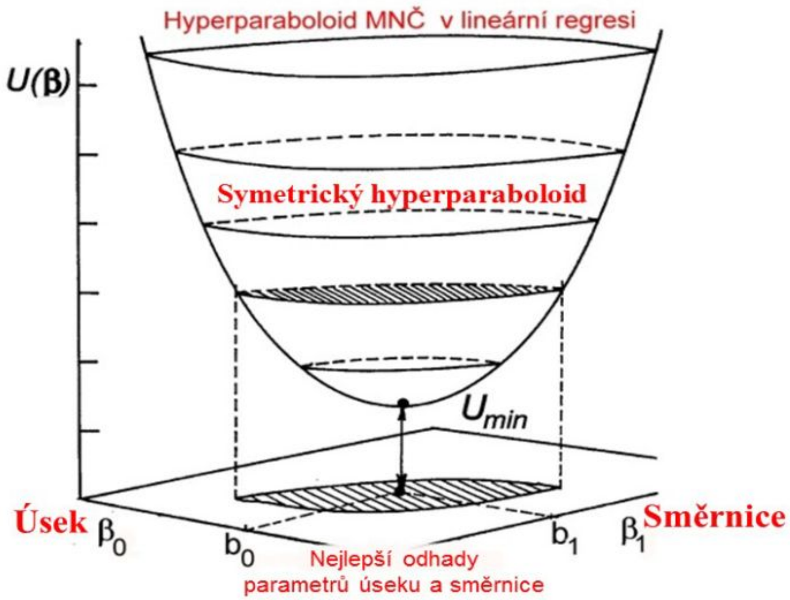
každou měřenou (experimentální) **hodnotu závisle proměnné** y_{exp}

hodnotou vypočtenou (predikovanou) y_{vyp}

čili **hodnotou** ležící na spojité funkci (modelu) **nezávisle proměnné** x .

Grafické vysvětlení regresního modelu:





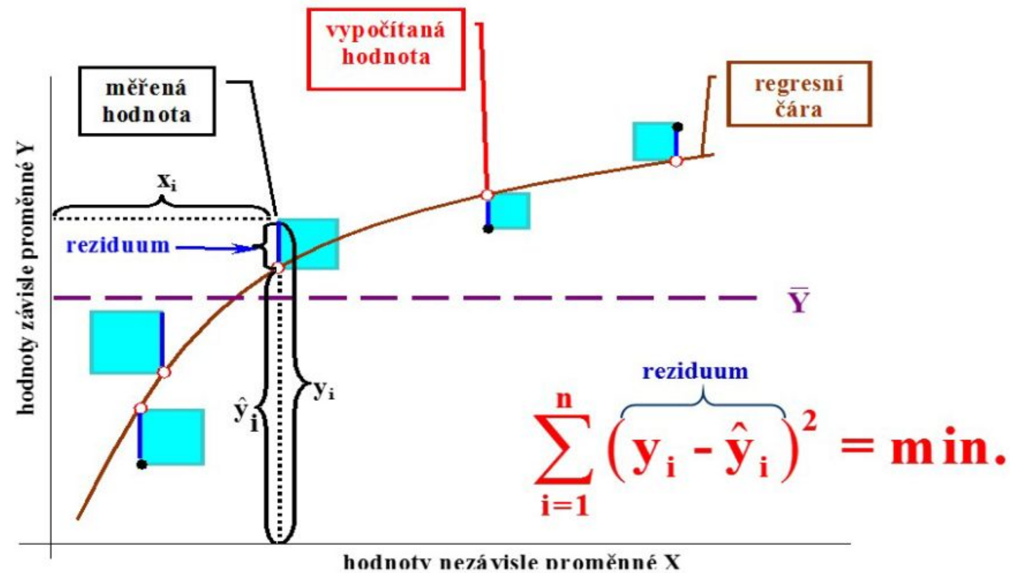
Účelová funkce U dosáhne minima pro nejlepší odhady parametrů úseku a směrnice

Formulace lineárního regresního modelu

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix}}_{\substack{\mathbf{y} \\ \text{závisle} \\ \text{proměnná}}} = \underbrace{\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{im} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nj} & \cdots & x_{nm} \end{bmatrix}}_{\substack{\mathbf{X} \\ \text{nezávisle proměnná}}} \cdot \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_j \\ \vdots \\ \beta_m \end{bmatrix}}_{\substack{\boldsymbol{\beta} \\ \text{regresní} \\ \text{parametry}}} + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{bmatrix}}_{\substack{\boldsymbol{\varepsilon} \\ \text{náhodná} \\ \text{chyba}}}$$

Maticový zápis $\mathbf{y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

Vyčíslení odhadů parametrů lineárního regresního modelu metodou nejmenších čtverců (MNČ)



Předpoklady metody nejmenších čtverců

1. Parametry $\boldsymbol{\beta}$ mohou nabývat libovolných hodnot. Omezení jsou pouze fyzikálního smyslu.
2. Model je lineární v parametrech $\boldsymbol{\beta}$; platí přitom additivní model měření $\mathbf{y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\epsilon}$.
3. Matice $\mathbf{X}$ je nenáhodná, nastavitelných hodnot nezávisle proměnných. Má hodnost m:
   (a) Žádné dva sloupce x_j a x_k nejsou kolineární (čili paralení).
   (b) $\mathbf{X}^T\mathbf{X}$ je pak symetrická regulární matice.
   (c) Rovina L je m rozměrná a vektory $\mathbf{X}\ \mathbf{b}$ jsou jednoznačné.
4. Náhodné chyby ϵ_i mají nulovou střední hodnotu $E(\epsilon_i) = 0$. Je-li $E(\epsilon_i) = K$, je nutno zavést absolutní člen a pak bude $E(\epsilon_i) = 0$.
5. Náhodné chyby ϵ_i mají konstantní rozptyl, homoskedasticita, $E(\epsilon_i^2) = \sigma^2$.
6. Náhodné chyby ϵ_i jsou vzájemně nekorelované, $\mathrm{cov}(\epsilon_i \cdot \epsilon_j) = E(\epsilon_i \cdot \epsilon_j) = 0$.
7. Náhodné chyby mají normální rozdělení $\epsilon \approx N(0, \sigma^2)$.

Koeficient determinace R^2 (×100%):
je čtverec vícenásobného korelačního koeficientu. Vyjadřuje procento bodů, popsaných regresním modelem.

Testování: H_0: $\boldsymbol{\beta}_c = 0$ nebo H_0: $R^2 = 0$

Testační kritérium F-testu

$$F_R = \frac{(CSC - RSC)(n - m)}{RSC(m - 1)} = \frac{\hat{R}^2 (n - m)}{(1 - \hat{R}^2)(m - 1)}$$

Je-li $F_R < F_{1-\alpha}(m - 1, n - m)$, je H_0 přijata.

Významnost korelačního koeficientu je shodná s testem významnosti všech regresních koeficientů vyjma absolutního členu.

Statistika	P608a	P608b	P608c	P608d
Úsek, b_0, s_0				
Směrnice b_1, s_1				
Test významnosti úseku, t_0				
Test významnosti směrnice, t_1				
Test celkové regrese, F_R				
Korelační koeficient, R				
Koeficient determinace, D				
Směrodatná odchylka, $s(y)$				
Trend v reziduích				
Závěr: model je				

Příklad 6.8 *Omezení klasické analýzy lineárního modelu*
Anscomb[5] uvádí testační data pro čtyři simulované výběry. Testujte statistickou významnost obou parametrů β_1 a β_2 a proveďte grafickou analýzu reziduí.
Data: čtyři simulované výběry vykazují stejné charakteristiky
$b_1 = 0.5$, $b_2 = 3.0$, $D(b_1) = 0.0139$ a $D(b_2) = 1.2656$.

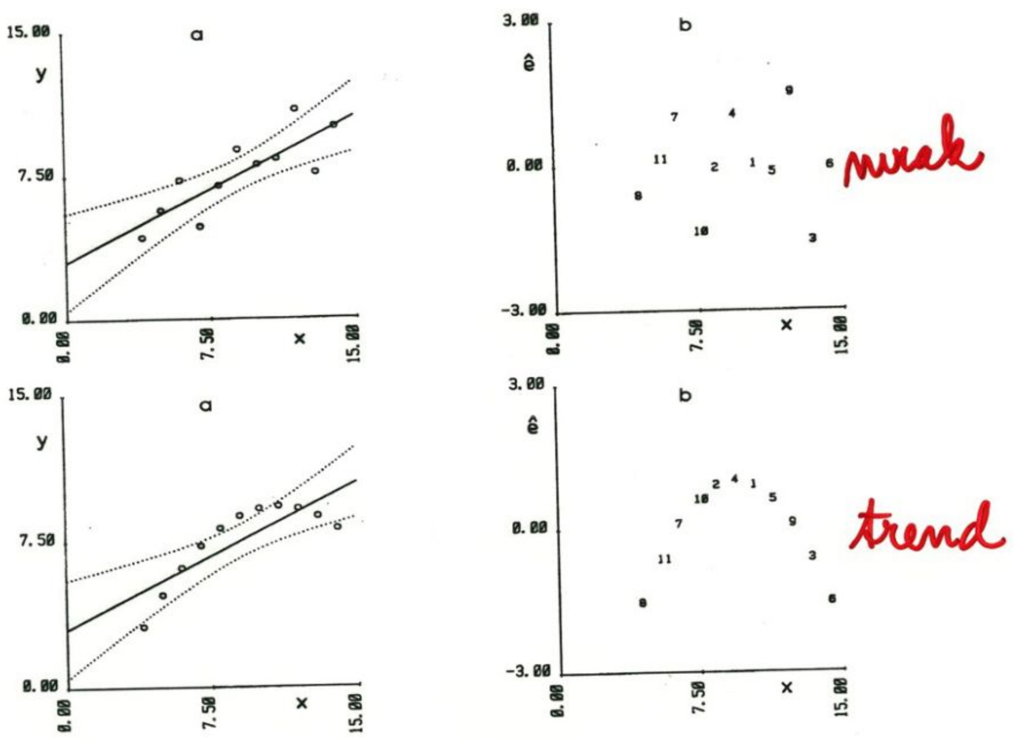
Výběr		A	B	C		D
Proměnná	x	y	y	y	x	y
1	10	8.04	9.14	7.46	8	6.58
2	8	6.95	8.14	6.77	8	5.76
3	13	7.58	8.74	12.74	8	7.71
4	9	8.81	8.77	7.11	8	8.84
5	11	8.33	9.26	7.81	8	8.47
6	14	9.96	8.10	8.84	8	7.04
7	6	7.24	6.13	6.08	8	5.25
8	4	4.26	3.10	5.39	19	12.50
9	12	10.84	9.13	8.15	8	5.56
10	7	4.82	7.26	6.42	8	7.91
11	5	5.68	4.74	5.73	8	6.89

Řešení:
Lineární regresní model $E(y/x) = \beta_1 x + \beta_2$ má pro všechny výběry

a) ***Stejné*** **odhady parametrů:**
$b_1 = 0.5$, $b_2 = 3.0$, $D(b_1) = 0.0139$, $D(b_2) = 1.2656$.

b) ***Stejné*** **testační statistiky významnosti parametrů:**
$T_1 = 2.667$ a $T_2 = 4.241$.

c) ***Stejné*** **testační statistiky:** $F_R = 17.97$, $\hat{R}^2 = 0.66$, $\hat{\sigma} = 1.237$ ukazují,
že β_1 a β_2 jsou významně odlišné od nuly.



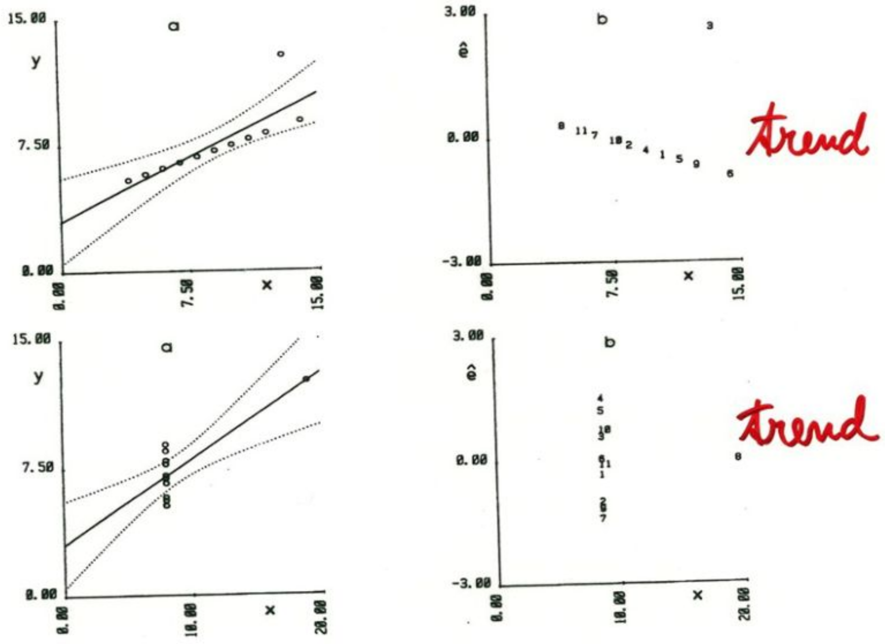
IS modelových hodnot přímky

Pro model přímky:

Směrodatná odchylka reziduí

$$\mu_{y'} = y_i' \underbrace{\pm t_{\frac{\alpha}{2},n-2}\cdot\frac{\sigma}{\sqrt{n-2}}\cdot\sqrt{1+\frac{n(x_i-\bar{x})^2}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}}}_{\text{Polovina IS modelu přímky}}$$

Modelová hodnota



Závěr: **Neshodu modelu s daty indikuje grafická analýza reziduí**

Intervalové odhady parametrů

Pro různý počet pozorování se mohou odhadnuté regresní parametry b_0 a b_1 lišit.

Vedle bodových odhadů regresních parametrů lze vyčíslit i jejich *intervalové odhady*:

$$b_i - t_{1-\alpha/2}(n-m)\cdot s(b_i) < \beta_i < b_i + t_{1-\alpha/2}(n-m)\cdot s(b_i)$$

kde b_i je bodový odhad regresního parametru,

$t_{1-\alpha/2}(n-p)$ je kvantil Studentova t rozdělení,

m je počet parametrů modelu,

$s(b_i)$ je směrodatná chyba odhadu parametru.

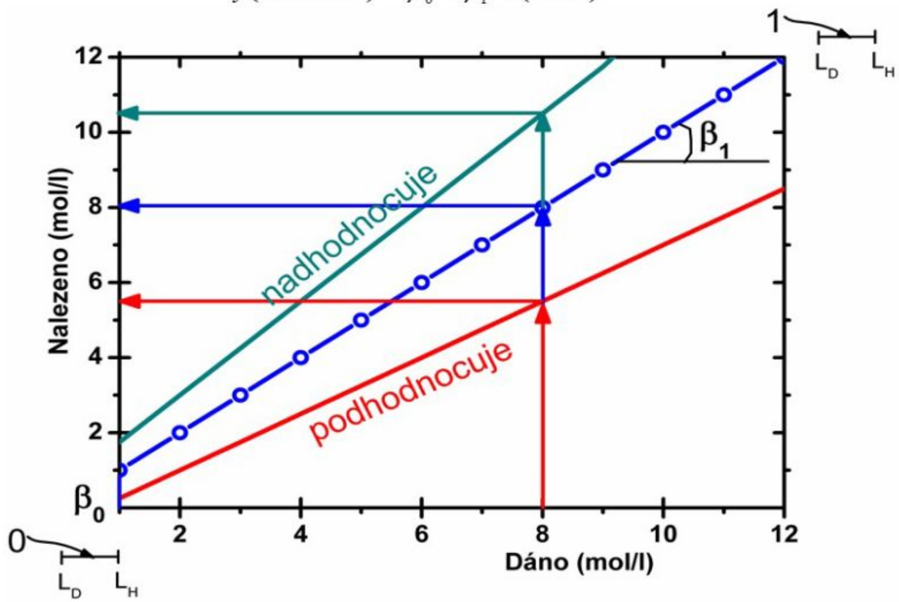
IS *y*-hodnot – Working-Hottelingův pás spolehlivosti

udává rozpětí, ve kterém se budou nacházet hodnoty závisle proměnné se zvolenou pravděpodobností $1 - \alpha$

$$y_{i(\min,\max)} = y_i' \pm t_{\frac{\alpha}{2};n-m} \cdot \sigma$$

Validace nové analytické metody

$y(\text{nalezeno}) = \beta_0 + \beta_1 \cdot x(\text{dáno})$



18

Příklad 6.7 *Validizace nové analytické metody*
Proveďte validizaci nové analytické metody porovnáním jejich výsledků y vůči standardům x .
(a) Určete odhady b_1 a b_2,
(b) zkonstruujte 95 %ní interval spolehlivosti úseku a směrnice,
(c) 95%ní elipsoid spolehlivosti,
(d) 95%ní interval spolehlivosti predikce v těžišti.
Data: n = 24, m = 2
Obsah látky určený standardní metodou (x) a novou metodou (y) v g:

x	40.2	43.8	47.6	50.7	56.8	81.3	83.3	97.1	102.5	118.7
y	48.9	39.1	42.6	56.9	70.3	71.5	97.6	99.9	105.2	102.3
x	129.4	184.8	287.5	295.4	420.3	421.3	427.9	566.1	608.5	640.7
y	106.8	162.9	234.0	303.4	388.8	391.1	369.3	611.6	580.2	643.3
	x		692.8		705.2		714.4		881.4	
	y		596.6		612.6		633.5		669.8	

Řešení:

(a) Odhady b_1 a b_2:
- ☐ odhad úseku $b_2 = 14.73$ (± 12.61),
- ☐ odhad směrnice $b_1 = 0.868$ (± 0.0302),
- ☐ koeficient determinace $\hat{R}^2 = 0.974$,
- ☐ odhad směrodatné odchylky reziduí $\hat{\sigma} = 39.54$.

(b) Interval spolehlivosti úseku: H_0: $\beta_2 = 0$ *vs.* H_A: $\beta_2 \neq 0$

$$b_2 - t_{1-\alpha/2}(22)\sqrt{D(b_2)} \leq \beta_2 \leq b_2 + t_{1-\alpha/2}(22)\sqrt{D(b_2)}$$

a dosazením

$$14.73 - 2.08 \times 12.61 \leq \beta_2 \leq 14.73 + 2.08 \times 12.61$$

vyjde

$$-11.499 \leq \beta_2 \leq 40.959$$

Závěr testování: interval spolehlivosti úseku zahrnuje nulu, takže H_0 je přijata a úsek β_2 lze považovat za nulový.

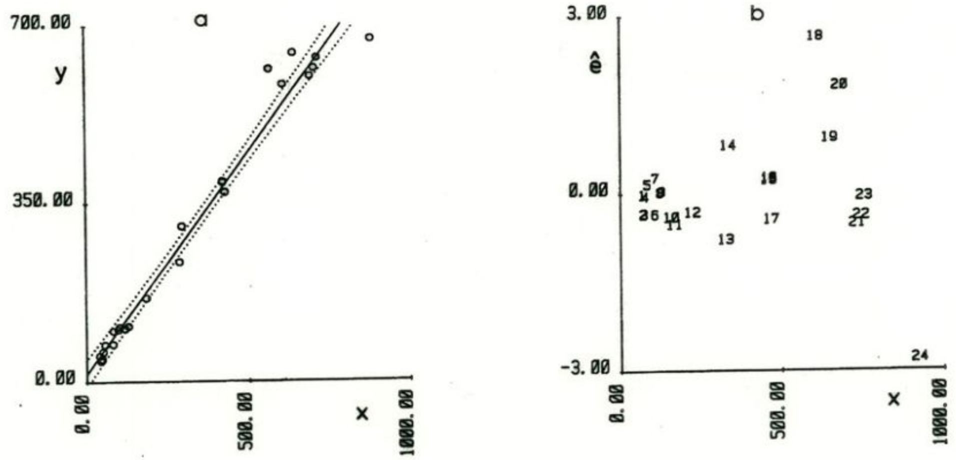
Interval spolehlivosti směrnice: H_0: $\beta_1 = 1$ *vs.* H_A: $\beta_1 \neq 1$

$$0.868 - 2.08 \times 0.0302 \leq \beta_1 \leq 0.868 + 2.08 \times 0.0302$$

a vyčíslením

$$0.805 \leq \beta_1 \leq 0.93$$

Závěr testování: interval spolehlivosti neobsahuje jedničku, takže H_0 je zamítnuta a směrnici β_1 nelze považovat za jednotkovou.



Závěr: intervaly spolehlivosti indikují, že úsek regresní přímky lze považovat za nulový $\beta_2 = 0$, zatímco směrnice β_1 je významně odlišná od jedničky.

Nová analytická metoda vede k odlišným výsledkům od standardní.

6.2.1 Úlohy na validaci nové analytické metody

Úloha V6.01 *Validace stanovení molybdenu rentg.-fluoresc. metodou*

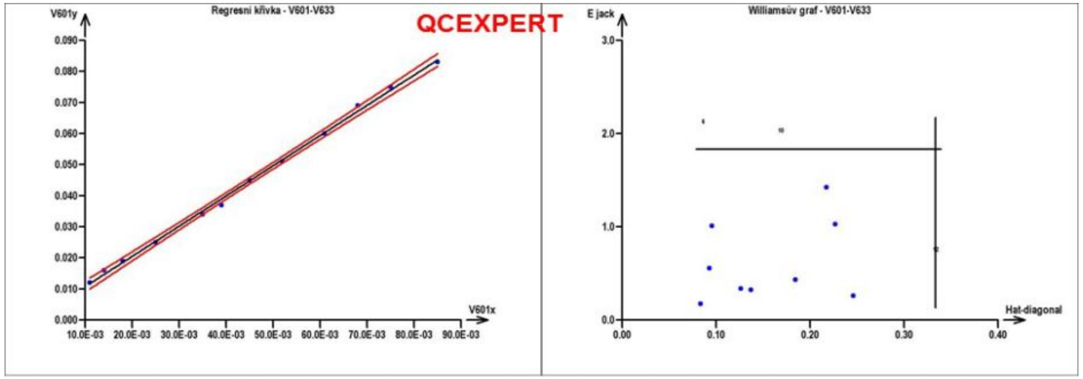
Zadání: U stanovení obsahu molybdenu porovnejte výsledky z rentg.-fluorescenční metody y s deklarovaným obsahem standardů ocelí x.

Úkoly:

(1) Určete velikost systematické chyby metody (= velikost úseku β_0).
(2) Správnost metody (= směrnice měla být 1).
(3) Pokuste se vyjádřit i přesnost metody.
(4) Jsou v datech vlivné a vybočující body?
(5) Tabulkové indikace vlivných bodů a pět nejdůležitějších grafů identifikace vlivných bodů.

Data: Obsah molybdenu, dáno x [%], stanoveno y [%]:

Dáno x	Stanoveno y
0.011	0.012
...	...
0.085	0.083



Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	0.001034	0.000686	Nevýznamný	0.163	-0.00049559	0.0025644
V601x	0.972702	0.013748	Významný	7.77E-015	0.9420701358	1.003335592

Statistické charakteristiky regrese

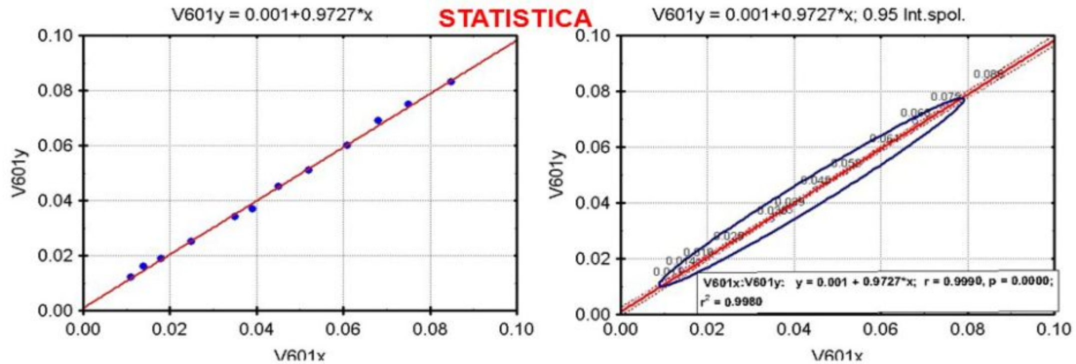
Vícenásobný korelační koeficient R :	0.99900
Koeficient determinace R^2 :	0.99800
Predikovaný korelační koeficient Rp :	0.99434

Střední kvdratická chyba predikce MEP : 1.50063E-006

Akaikeho informační kritérium : -161.13

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F :	5005.80
Kvantil F (1-alfa, m-1, n-m) :	4.96460
Pravděpodobnost :	7.75E-015
Závěr :	Model je významný



Efekt	Odhady parametrů (V6.sta) Sigma-omezená parametrizace V601y Param.	V601y Sm.Ch.	V601y t	V601y p	-95.00% LmtSpol.	+95.00% LmtSpol.	V601y Beta (ß)	V601y Sm.Ch. ß
Abs. člen	0.001034	0.000687	1.50641	0.162882	-0.000496	0.002564		
"V601x"	*0.972703*	*0.013748*	*70.75168*	*0.000000*	*0.942070*	*1.003336*	*0.999003*	*0.014120*

NCSS2007

Linear Regression Plot Section

Run Summary Section

Parameter	Value	Parameter		Value
Dependent Variable		V601y	Rows Processed	81
Independent Variable		V601x	Rows Used in Estimation	12
Frequency Variable		None	Rows with X Missing	69
Weight Variable		None	Rows with Freq Missing	0
Intercept	0.0010		Rows Prediction Only	0
Slope	0.9727		Sum of Frequencies	12
R-Squared	0.9980		Sum of Weights	12.0000
Correlation	0.9990		Coefficient of Variation	0.0257
Mean Square Error	1.267129E-06		Square Root of MSE	1.125668E-03

Summary Statement

The equation of the straight line relating V601y and V601x is estimated as: V601y = (0.0010) +(0.9727) V601x using the 12 observations in this dataset. The y-intercept, the estimated value of V601y when V601x is zero, is 0.0010 with a standard error of 0.0007. The slope, the estimated change in V601y per unit change in V601x, is 0.9727 with a standard error of 0.0137. The value of R-Squared, the proportion of the variation in V601y that can be accounted for by variation in V601x, is 0.9980. The correlation between V601y and V601x is 0.9990. A significance test that the slope is zero resulted in a t-value of 70.7517. The signifikance level of this t-test is 0.0000. Since 0.0000 < 0.0500, the hypothesis that the slope is zero is rejected. The estimated slope is 0.9727. The lower limit of the 95% confidence interval for the slope is 0.9421 and the upper limit is 1.0033. The estimated intercept is 0.0010. The lower limit of the 95% confidence interval for the intercept is -0.0005 and the upper limit is 0.0026.

Descriptive Statistics Section

Parameter	Dependent	Independent
Variable	V601y	V601x
Count	12	12
Mean	0.0438	0.0440
Standard Deviation	0.0240	0.0247
Minimum	0.0120	0.0110
Maximum	0.0830	0.0850

Regression Estimation Section

Parameter	Intercept B(0)	Slope B(1)
Regression Coefficients	0.0010	0.9727
Lower 95% Confidence Limit	-0.0005	0.9421
Upper 95% Confidence Limit	0.0026	1.0033
Standard Error	0.0007	0.0137
Standardized Coefficient	0.0000	0.9990
T Value	1.5064	70.7517
Prob Level (T Test)	0.1629	0.0000
Reject H0 (Alpha = 0.0500)	No	Yes
Power (Alpha = 0.0500)	0.2759	1.0000
Regression of Y on X	0.0010	0.9727
Inverse Regression from X on Y	0.0009	0.9746
Orthogonal Regression of Y and X	0.0010	0.9736

Notes:

The above report shows the least-squares estimates of the intercept and slope followed by the corresponding standard errors, confidence intervals, and hypothesis tests. Note that these results are based on several assumptions that should be validated before they are used.

Estimated Model: (1.03440731901351E-03) + (.972702863961814) * (V601x)

Correlation and R-Squared Section

Parameter	Pearson Correlation Coefficient	R-Squared	Spearman Rank Correlation Coefficient
Estimated Value	0.9990	0.9980	1.0000
Lower 95% Conf. Limit (r dist'n)	0.9960		
Upper 95% Conf. Limit (r dist'n)	0.9995		
Lower 95% Conf. Limit (Fisher's z)	0.9963		1.0000
Upper 95% Conf. Limit (Fisher's z)	0.9997		1.0000
Adjusted (Rbar)	0.9978		
T-Value for H0: Rho = 0	70.7517	70.7517	
Prob Level for H0: Rho = 0	0.0000	0.0000	0.0000

Notes:

The confidence interval for the Pearson correlation assumes that X and Y follow the bivariate normal distribution. This is a different assumption from linear regression which assumes that X is fixed and Y is normally distributed. Two confidence intervals are given. The first is based on the exact distribution of Pearson's correlation. The second is based on Fisher's z transformation which approximates the exact distribution using the normal distribution. Why are both provided? Because most books only mention Fisher's approximate method, it will often be needed to do homework. However, the exact methods should be used whenever possible. The confidence limits can be used to test hypotheses about the correlation. To test the hypothesis that rho is a specific value, say r0, check to see if r0 is between the confidence limits. If it is, the null hypothesis that rho = r0 is not rejected. If r0 is outside the limits, the null hypothesis is rejected. Spearman's Rank correlation is calculated by replacing the orginal data with their ranks.
This correlation is used when some of the assumptions may be invalid.

Tests of Assumptions Section

Assumption/Test	Test Value	Prob Level	Is the Assumption Reasonable at the 0.2000 Level of Significance?
Residuals follow Normal Distribution?			
Shapiro Wilk	0.9853	0.996849	Yes
Anderson Darling	0.1507	0.962228	Yes
D'Agostino Skewness	0.0094	0.992478	Yes
D'Agostino Kurtosis	0.0319	0.974562	Yes
D'Agostino Omnibus	0.0011	0.999447	Yes
Constant Residual Variance?			
Modified Levene Test	0.1117	0.745133	Yes
Relationship is a Straight Line?			
Lack of Linear Fit F(0, 0) Test	0.0000	0.000000	No

No Serial Correlation?

Evaluate the Serial-Correlation report and the Durbin-Watson test if you have equal-spaced, time series data.

Notes:

A 'Yes' means there is not enough evidence to make this assumption seem unreasonable. This lack of evidence may be because the sample size is too small, the assumptions of the test itself are not met, or the assumption is valid. A 'No' means the that the assumption is not reasonable. However, since these tests are related to sample size, you should assess the role of sample size in the tests by also evaluating the appropriate plots and graphs. A large dataset (say N > 500) will often fail at least one of the normality tests because it is hard to find a large dataset that is perfectly normal.

Normality and Constant Residual Variance:

Possible remedies for the failure of these assumptions include using a transformation of Y such as the log or square root, correcting data-recording errors found by looking into outliers, adding additional independent variables, using robust regression, or using bootstrap methods.

Straight-Line: Possible remedies for the failure of this assumption include using nonlinear regression or polynomial regression.

Úloha V6.02 *Bichromátometrická metoda stanovení železitých iontů*

Zadání: Kraft a Dosch[60] navrhli titrační stanovení železa ve vodách. Železité ionty Fe^{3+} v Fe_2O_3 se redukují titanitou solí v přebytku a vzniklé ionty Fe^{2+} se pak stanoví bichromátometricky.

Úkoly:

(1) Vede titrační stanovení ke správným výsledkům?

(2) Proveďte Studentův *t*-test významnosti úseku b_0 (má být $\beta_0 = 0$).

(3) Proveďte Studentův *t*-test jednotkové směrnice b_1 (má být $\beta_1 = 1$).

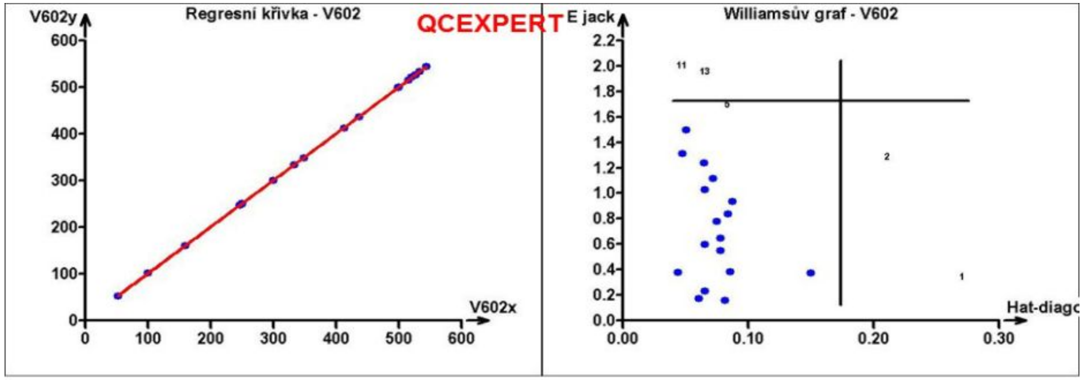
(4) Proveďte kombinovaný test obou parametrů v modelu přímky.

(5) Popište test významnosti absolutního členu.

(6) Popište test vhodnosti lineárního modelu dle Uttsové.

Data: Obsah Fe_2O_3 [mg], dáno *x*, nalezeno *y*:

Dáno *x*	Stanoveno *y*
52.0	52.50
...	...
543.61	543.78



Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	0.70845	0.23872	Významný	0.007343	0.21200	1.20490
V602x	0.99834	0.00056	Významný	0	0.99716	0.99951

Statistické charakteristiky regrese

Vícenásobný korelační koeficient R :	0.99999
Koeficient determinace R^2 :	0.99999
Predikovaný korelační koeficient Rp :	0.99998
Střední kvdratická chyba predikce MEP :	0.17699
Akaikeho informační kritérium :	-39.545

Úloha V6.04 *Stanovení kyseliny ftalové tenkovrstvou chromatografií*

Zadání: Obsah kyseliny ftalové byl stanoven tenkovrstvou chromatografií a chromatogram byl vyhodnocován remisním fotometrem.

Úkoly:

(1) Stanovte oba parametry lineárního regresního modelu a vyšetřete, zda je úsek nulový a směrnice jednotková.

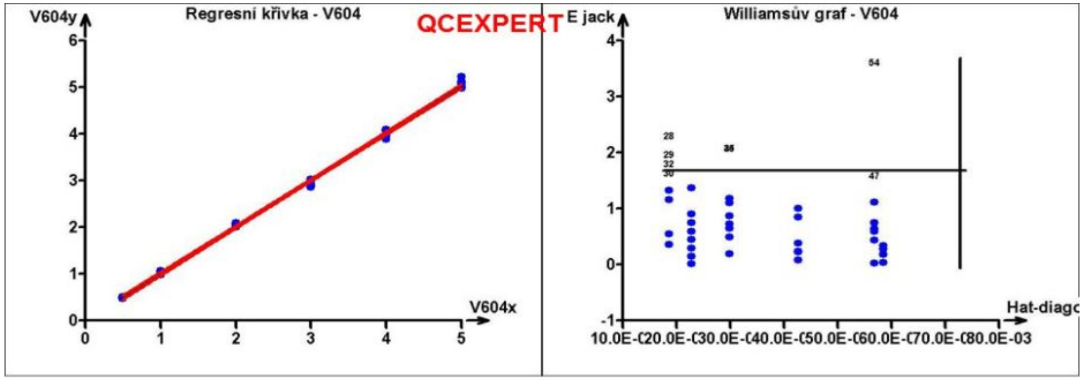
(2) Vyšetřete, zda jsou v datech vybočující hodnoty?

(3) Je stanovení je správné?

(4) Jaký je nutno zvolit postup při porušení předpokladů MNČ?

Data: Obsah kyseliny ftalové [μg], dáno *x*, nalezeno *y* (opakovaně).

Dáno *x*	Stanoveno *y*
0.50	0.48
...	...
5.23	65.02



Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	-0.0110	0.0187	Nevýznamný	0.5565	-0.048499	0.026400
V604x	1.00588	0.0059	Významný	0	0.9940574	1.017716

Statistické charakteristiky regrese

Vícenásobný korelační koeficient R :	0.9990902542
Koeficient determinace R^2 :	0.9981813361
Predikovaný korelační koeficient Rp :	0.9961043053
Střední kvdratická chyba predikce MEP :	0.004614241291
Akaikeho informační kritérium :	-295.6513242

Úloha V6.06 *Ověření stanovení železa spektrofotometrickou metodou*

Zadání: Ověřte stanovení obsahu železa *y* v $CoSO_4$ spektrofotometricky SFM *y* porovnáním výsledků standardního stanovení obsahu *x* metodou AAS, u které je předpokládána zanedbatelná náhodná chyba.

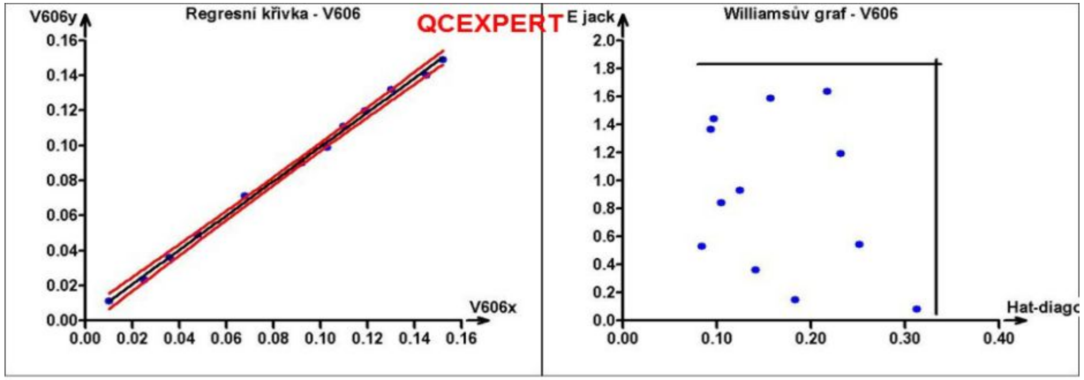
Úkoly:

(1) Vedou obě metody ke shodným výsledkům?

(2) Jsou v datech odlehlé hodnoty? Užijte pět grafů indikace vlivných bodů.

Data: Obsah železa v $CoSO_4$ [%], když je AAS *x* [%], SFM *y* [%]:

Dáno *x*	Stanoveno *y*
0.010	0.011
...	...
0.152	0.149



Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	0.001011	0.001531	Nevýznamný	0.5238	-0.0023997	0.0044219
V606x	0.981567	0.015615	Významný	2.5313E-014	0.9467727	1.0163614

Statistické charakteristiky regrese

Vicenásobný korelační koeficient R :	0.9987368986
Koeficient determinace R^2 :	0.9974753927
Predikovaný korelační koeficient Rp :	0.9929318535
Střední kvdratická chyba predikce MEP :	7.273899498E-006
Akaikeho informační kritérium :	-142.032264

Úloha V6.07 *Ověření stanovení dusičnanů v pitné a povrchové vodě*

Zadání: V chemických laboratořích geochemické firmy se zavedla nová metoda stanovení obsahu dusičnanů y v pitných ale také povrchových vodách pomocí iontově párové chromatografie.

Úkoly:

(1) Validujte novou metodu vůči deklarovaným obsahům NO_3^- [mg/l] *x*.

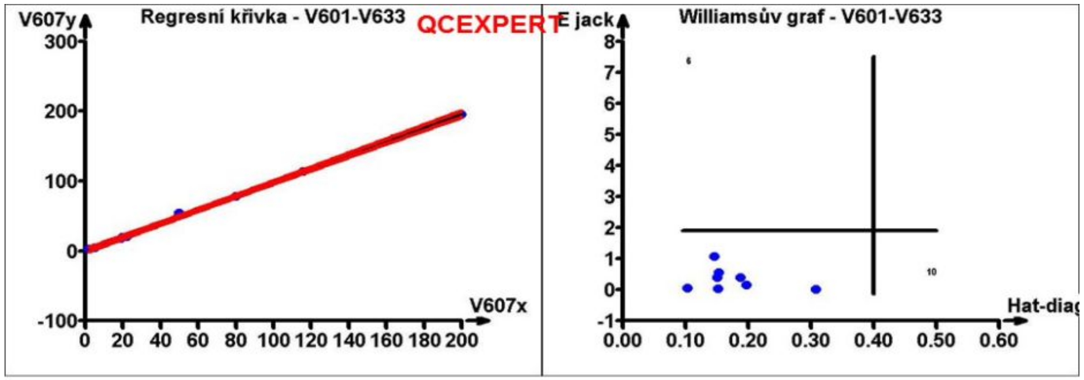
(2) Odhadněte regresní parametry metodu ortogonální regrese.

(3) Vede nová metoda ke správným výsledkům?

(4) Proveďte simultánní test významnosti úseku a významnosti směrnice, zda je rovna jedné.

Data: Pro obsah dusičnanů NO_3^- [mg/l] je dáno *x*, nalezeno *y*.

Dáno *x*	Stanoveno *y*
2.10	2.20
...	...
200.00	195.00



Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	-0.14284	0.9389	Nevýznamný	0.88284	-2.3079	2.0222
V607x	0.9801479567	0.0098	Významný	1.179E-013	0.95739	1.0029

Statistické charakteristiky regrese

Vicenásobný korelační koeficient R :	0.9995948631
Koeficient determinace R^2 :	0.9991898903
Predikovaný korelační koeficient Rp :	0.9978300448
Střední kvdratická chyba predikce MEP :	4.643796689
Akaikeho informační kritérium :	16.4284433

Úloha V6.20 *Validace nové metody stanovení arsenu v odpadní vodě*

Zadání: Je třeba validovat nové jednodušší stanovení arsenu v odpadní vodě. Mezi naměřenou koncentrací arsenu y a známou koncentrací x v µg/ml je předpokládán lineární regresní model $y = \beta_0 + \beta_1 x$.

Úkoly:

(1) Užitím ortogonální regrese ověřte správnost nové metody.

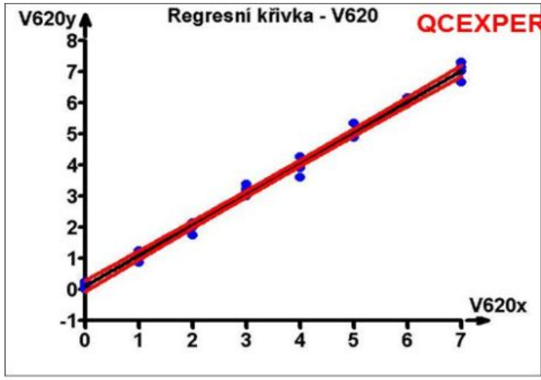
(2) K jakému výsledku dospěje nová metoda, když standard arsen vůbec neobsahuje čili absolutní člen je nulový, $\beta_0 = 0$?

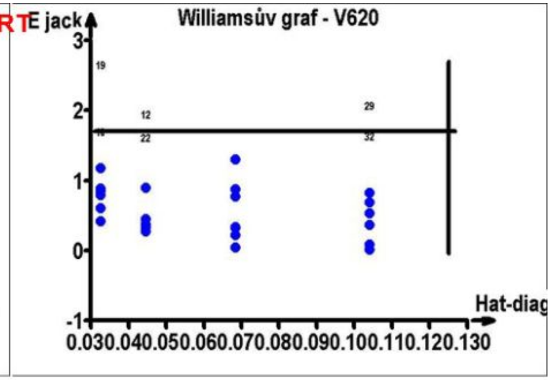
(3) Vyšetřete, zda nová metoda nadhodnocuje či podhodnocuje?

(4) Jakou modifikaci MNČ je třeba použít, když jsou všechny proměnné zatíženy náhodnými chybami?

Data: Koncentrace arsenu daná x [µg. cm^{-3}], nalezená y [µg. cm^{-3}].

Dáno x	Stanoveno y
0	0.17
...	...
7.0	7.30





Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	0.10458	0.06051	Nevýznamný	0.0942	-0.01899	0.228167
V620x	0.98770	0.01446	Významný	0	0.958168	1.017248

Statistické charakteristiky regrese

Vícenásobný korelační koeficient R :	0.99679
Koeficient determinace R^2 :	0.99360
Predikovaný korelační koeficient Rp :	0.98563
Střední kvdratická chyba predikce MEP :	0.03715
Akaikeho informační kritérium :	-105.20

Úloha V6.22 *Validace navržené titrační metody ke stanovení modré báze MB H-3R*

Zadání: Při výrobě modré báze MB H-3R byl stanovován její obsah v pastě z kalolisu titračně dusitanem v kyselém prostředí y a standardně spektrofotometricky x. Za základ byla vzata titrační metoda. Rozptyl této metody se považuje za zanedbatelný vůči rozptylu spektrofotometrické metody.

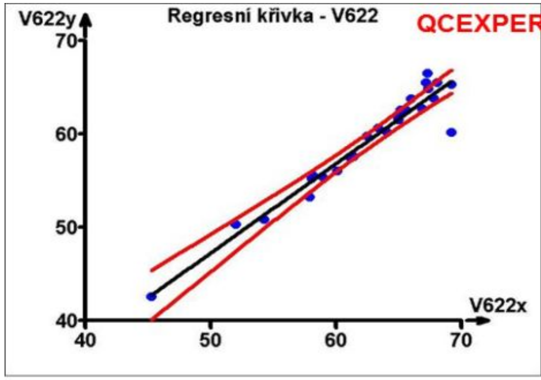
Úkoly:

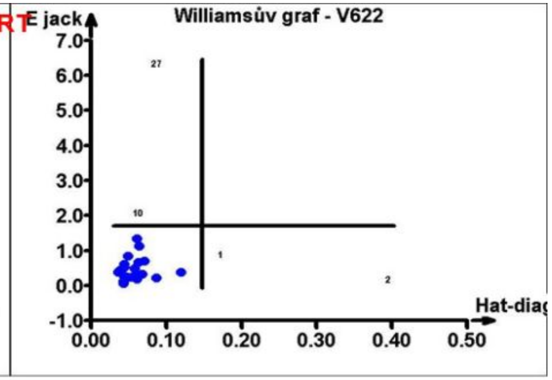
(1) Popište test významnosti absolutního členu.

(2) Vysvětlete test shodnosti odhadu parametru $\boldsymbol{\beta}$ s předepsanou $\boldsymbol{\beta}_0$.

Data: Koncentrace modré báze spektrofotometrickou metodou x a titrační metodou y.

Dáno x	Stanoveno y
52.0	50.3
...	...
69.2	60.1





Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	-0.65650	3.14318	Nevýznamný	0.83624	-7.1300	5.8170
V622x	0.957240	0.04994	Významný	2.220E-016	0.85438	1.0600

Statistické charakteristiky regrese

Vícenásobný korelační koeficient R :	0.967619281
Koeficient determinace R^2 : 0.936287073	
Predikovaný korelační koeficient Rp :	0.8551935541
Střední kvdratická chyba predikce MEP :	2.300430311
Akaikeho informační kritérium :	22.00604283

Úloha V6.31 *Validace stanovení chromu metodou AAS a ICP-AES*

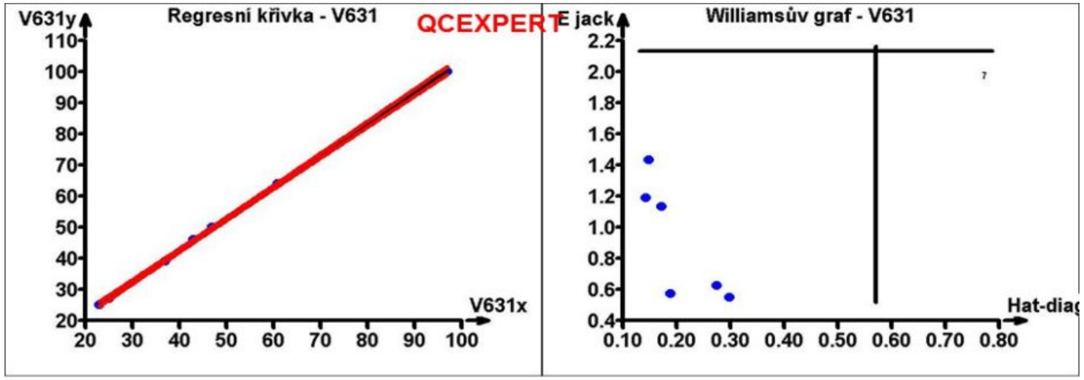
Zadání: Ve vzorcích půdy byl stanoven metodami AAS a ICP-AES obsah chromu.

Úkoly:

(1) Porovnejte shodnost výsledků stanovení oběma metodami.
(2) Vysvětlete 7 předpokladů MNČ a řešení regresního tripletu.
(3) Ukažte postup validace nové analytické metody testování nulovosti úseku a jednotkovosti směrnice.
(4) Jak se bude řešit tato úloha v případě porušení předpokladů MNČ?

Data: *x* značí AAS [mg/kg], *y* značí ICP-AES [mg/kg]:

Dáno *x*	Stanoveno *y*
25	27
...	...
97	100



Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	1.86193	0.35244	Významný	0.00323	0.95595	2.76791
V631x	1.01491	0.00664	Významný	2.27699E-10	0.99784	1.03198

Statistické charakteristiky regrese

Vicenásobný korelační koeficient R :	0.9998929533
Koeficient determinace R^2 : 0.999785918	
Predikovaný korelační koeficient Rp :	0.9985017625
Střední kvdratická chyba predikce MEP :	0.4268212711
Akaikeho informační kritérium :	-10.73012051