RG1) Proveďte zobecněnou lineární regresi následujících dat

x_i	-2.8	-2.4	-2.0	-1.6	-1.2	-0.8	-0.4	0
y_i	-13.5	-6.0	1.1	6.0	6.1	6.9	4.1	3.5
x_i	0.4	0.8	1.2	1.6	2.0	2.4	2.8	—
y_i	1.2	-0.7	0.1	-2.5	2.3	4.3	1.1	—

polynomem za předpokladu, že chyba měření je konstantní. Vyberte nejvhodnější stupeň polynomu, určete neznámé koeficienty polynomu a jejich směrodatné odchylky. Odhadněte chybu měření a pomocí znaménkového testu zkontrolujte vhodnost modelu.

RG2) Proveďte zobecněnou lineární regresi následujících dat

x_i	0	0.45	0.9	1.35	1.79	2.24	2.69	3.14
y_i	2.74	1.70	1.57	0.04	-1.30	-0.70	-1.36	-0.78
x_i	3.59	4.04	4.49	4.93	5.38	5.83	6.28	—
y_i	-0.50	-0.31	1.63	2.74	2.78	3.77	1.39	—

pomocí rozvoje do harmonických funkcí

$$y = a_0 + a_1 \cos \frac{2\pi(x - x_{min})}{x_{max} - x_{min}} + b_1 \sin \frac{2\pi(x - x_{min})}{x_{max} - x_{min}}$$

za předpokladu, že chyba měření je konstantní. Určete neznámé koeficienty a_0, a_1, b_1 a jejich směrodatné odchylky. Odhadněte chybu měření a pomocí znaménkového testu zkontrolujte vhodnost modelu. Uvedený model porovnejte s modelem obsahujícím navíc ještě členy s dvojnásobnou frekvencí.

RG3) Aproximujte následující data

x_i	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
y_i	9.21	7.33	5.30	4.09	2.90	2.20	1.82	1.13
x_i	0.8	0.9	1	1.1	1.2	1.3	1.4	—
y_i	0.947	0.825	0.440	0.399	0.193	0.176	0.221	—

vzorcem $y(x) = a\exp(-bx)$ za předpokladu, že chyba měření $\delta y = C\sqrt{y}$, kde C je neznámá konstanta. Úlohu převeďte na lineární regresi. Určete neznámé parametry a, b a jejich směrodatné odchylky. Odhadněte konstantu C. Proveďte i výpočet za předpokladu konstantní odchylky měření a výsledky obou výpočtů porovnejte.

RG4) Aproximujte následující data

x_i	-1.4	-1.2	-1.0	-0.8	-0.6	-0.4	-0.2	0
y_i	0.06	-0.077	0.30	0.315	0.596	0.882	0.776	0.846
x_i	0.2	0.4	0.6	0.8	1.0	1.2	1.4	—
y_i	0.905	0.682	0.477	0.249	0.187	0.162	0.034	—

vzorcem $y(x) = a\exp(-bx^2)$ za předpokladu, že chyba měření $\delta y = C\sqrt{y}$, kde C je neznámá konstanta. Úlohu můžete převést na zobecněnou lineární regresi. Určete neznámé parametry a, b a jejich směrodatné odchylky. Odhadněte konstantu C. Proveďte i výpočet za předpokladu konstantní odchylky měření a výsledky obou výpočtů porovnejte.

RG5) Aproximujte následující data

x_i	-1.4	-1.2	-1.0	-0.8	-0.6	-0.4	-0.2	0
y_i	0.281	0.549	0.622	0.878	1.121	0.937	0.776	0.632
x_i	0.2	0.4	0.6	0.8	1.0	1.2	1.4	—
y_i	0.905	0.682	0.477	0.249	0.187	0.162	0.034	—

vzorcem $y(x) = a\exp\left[-b(x-c)^2\right]$ za předpokladu, že chyba měření $\delta y = Cy$, kde C je neznámá konstanta. Úlohu můžete převést na zobecněnou lineární regresi tak, že $\ln(y)$ aproximujete kvadratickým polynomem. Určete neznámé parametry a, b, c a jejich směrodatné odchylky. Odhadněte konstantu C.

RG6) Proveďte zobecněnou lineární regresi následujících dat

x_i	0.5	1.0	1.5	2.0	2.5	3.0
y_i	7.5	4.5	3.0	4.0	4.0	5.0
x_i	3.0	4.0	5.0	6.0	—	—
y_i	4.0	5.0	6.5	7.5	—	—

funkcí $y(x) = a_1 + a_2\, x + a_3\, /\, x$ za předpokladu konstantních neznámých chyb měření, odhadněte chybu měření, stanovte směrodatné odchylky a kovariance koeficientů a_1, a_2, a_3. Nakreslete graf reziduí a odhadněte přípustnost modelu.

RG7) Proveďte zobecněnou lineární regresi následujících dat

x_i	0.5	1.0	1.5	2.0	2.5	3.0
y_i	7.5	4.5	3.0	4.0	4.0	5.0
x_i	3.0	4.0	5.0	6.0	—	—
y_i	4.0	5.0	6.5	7.5	—	—

funkcí $y(x) = a_1 + a_2\, x^2 + a_3\, /\, x^2$ za předpokladu konstantních neznámých chyb měření, odhadněte chybu měření, stanovte směrodatné odchylky a kovariance koeficientů a_1, a_2, a_3. Nakreslete graf reziduí a odhadněte přípustnost modelu.

RG8) Proveďte aproximaci následujících dat

x_i	0.5	1.0	1.5	2.0	2.5	3.0
y_i	0.5	1.2	2.7	1.8	3.0	3.3
x_i	3.5	4.0	5.0	6.0	—	—
y_i	3.6	4.1	3.7	5.9	—	—

pomocí lineární závislosti $y = a + b \cdot x$ metodou nejmenších čtverců a metodou minimalizace sumy absolutních hodnot reziduí. Proveďte analýzu výsledku regrese pomocí metody nejmenších čtverců. Porovnejte oba přístupy.

RG9) Nastudujte podle knihy Numerical Recipes postup regrese přímkou v případě, kdy chyby měření nezávisle proměnné nelze zanedbat. Vyzkoušejte procedury z knihy při prosté lineární regresi následujících dat

x_i	0.5	1.0	1.5	2.0	2.5	3.0
y_i	0.5	1.2	1.7	1.8	3.0	3.3
x_i	3.5	4.0	5.0	6.0	—	—
y_i	3.6	4.1	4.3	5.9	—	—

Předpokládejte, chyby měření x a y jsou stejné a konstantní.

RG10) Nagenerujte data

$$y(x) = \exp(-x) + 0.1\, a(x) \sqrt{\exp(-x)}$$

v bodech $x_i = 0.02k$, kde k = 0,. . . , 200. Čísla $a(x_i)$ jsou navzájem nezávislá náhodná čísla s normálním rozdělením $N(0, 1)$. Nalezněte koeficienty α, β modelu $y(x) = \alpha \exp(-\beta x)$, můžete provést. Odhadněte přesnost stanovení koeficientů a regresi vyhodnoťte.

RG11) Pro $x_i = (i-1) * 0.1$, kde $i = 1, \ldots, 100$ napočtěte data $y_i = -0.67 + 0.35 x_i + e_i$, kde e_i je náhodné číslo s rozdělením $f(x) = 0.5 \exp(-|x|)$. Z takto vygenerovaných dat odhadněte koeficienty lineární závislosti metodou nejmenších čtverců a metodou minimalizace sumy absolutních hodnot reziduí.

RG12) Pro následující data

x_1	x_2	x_3	x_4	y		x_1	x_2	x_3	x_4	y
217	67	260	91	481		141	52	190	66	292
152	58	203	68	338		153	56	183	70	357
180	66	170	77	396		193	71	178	82	429
162	65	160	74	345		180	80	170	84	469
205	77	188	83	425		168	74	170	79	358
232	65	220	72	393		146	68	158	68	346
173	51	243	56	279		155	64	198	59	311
212	66	220	77	401		138	70	180	62	267
147	54	150	75	404		197	76	228	88	442
165	59	188	70	368		125	58	160	66	295
161	52	190	69	391		132	62	163	59	264
257	64	313	96	487		236	72	225	84	481
149	57	173	68	374		161	57	173	65	309
198	59	220	62	367		245	70	218	69	469
141	63	193	75	396		177	53	183	75	338

proveďte vícenásobnou lineární regresi, vyhodnoťte významnost koeficientů a dále uvažujte jen ty proměnné, na kterých y významně závisí. Proveďte a vyhodnoťte regresi.

RG13) Pomocí nelineární regrese určete parametry aproximace následujících dat

x_i	10	11	12	13	14	15	16	17	18	19
y_i	0	2	0	4	8	37	39	43	19	24
x_i	20	21	22	23	24	25	26	27	28	29
y_i	21	18	40	51	63	72	57	64	38	35
x_i	30	31	32	33	34	35	36	37	38	39
y_i	20	18	7	5	4	2	2	3	1	2

pomocí sumy 2 normálních rozdělení s neznámými parametry

$$y = \sum_{i=1}^{2} a_i \exp\left[-\frac{(x-b_i)^2}{c_i^2}\right]$$

za předpokladu, že chyba měření je konstantní. Určete neznámé koeficienty a_i, b_i, c_i a jejich směrodatné odchylky. Můžete využít některého statistického programu, např. SYSTAT (poskytnu).

RG14) Pomocí nelineární regrese určete parametry aproximace dat z příkladu RG13 sumou dvou Lorentzových rozdělení

$$y = \sum_{i=1}^{2} \frac{a_i}{b_i^2 + (x-c_i)^2}$$

za předpokladu, že chyba měření je konstantní. Určete neznámé koeficienty a_i, b_i, c_i a jejich směrodatné odchylky. Můžete využít některého statistického programu, např. SYSTAT (poskytnu).

RG15) Vygenerujte data $y_i = (5 + 0.1\, x_i^2)(1 + e_i)$ pro $x_i = 0.25i,\ i = 1, \ldots, 20$ a e_i je náhodné číslo s rovnoměrným rozdělením v intervalu $\langle -0.25, 0.25)$. Proveďte zobecněnou lineární regresi polynomem 2.řádu, regresi řádně vyhodnoťte.

RG16) Vygenerujte data y_i s Poissonovým rozdělením (např. funkce POIDEV z Numerical Recipes) s parametrem $\lambda_i = 10\exp(-0.1 * i)$ pro $i = 1, \ldots, 20$. Pomocí regrese určete parametry modelu $\lambda = a\ \exp(-bt)$. Model můžete převést na lineární.

HY1) Prověřte předpoklad, že následující soubor dat má střední hodnotu $\overline{x} = 3$ a medián $x_m = 3$,

3.15E+00	4.40E+00	2.34E+00	2.22E+00	3.08E+00	3.71E+00	4.11E+00
4.04E+00	4.27E+00	4.19E+00	2.20E+00	3.81E+00	4.18E+00	6.78E+00
3.54E+00	2.20E+00	3.76E+00	2.58E+00	3.73E+00	2.48E+00	

HY2) Prověřte předpoklad, že soubor dat z příkladu HY1) směrodatnou odchylku $\sigma = 0.5$ a normální rozdělení.

HY3) Prověřte předpoklady, že vzájemně nekorelované soubory dat x_i a y_i

x_i	4.04E+00	3.43E+00	7.29E-01	1.64E+00	2.32E+00
x_i	3.77E+00	1.87E+00	2.22E+00	2.64E+00	4.54E+00
y_i	2.35+00	2.33+00	4.29+00	4.14+00	3.82+00
y_i	4.66+00	2.36+00	5.62+00	4.26+00	3.57+00

mají shodné mediány a rozdělení.

HY4) Prověřte předpoklady, že vzájemně nekorelované soubory dat x_i a y_i z příkladu HY3) mají shodné střední hodnoty a směrodatné odchylky.

HY5) Určete lineární korelační koeficient a Spearmannův koeficient korelace pořadí pro 2 binárně korelované soubory dat (např. x_i, y_i jsou 2 veličiny změřené na i-tém vzorku). Prověřte pomocí korelace hodnot i korelace pořadí hypotézu o nekorelovanosti veličin x a y. Data jsou shrnuta v tabulce

x_i	3.721	2.043	3.114	6.599	4.896	3.996
y_i	4.505	6.836	5.644	4.947	6.314	6.472
x_i	5.828	3.870	4.727	6.100	4.238	4.085
y_i	8.161	6.464	6.823	3.994	6.607	6.945

HY6) Při hodnocení životnosti stroje byly zjištěny tyto doby do poruchy 16, 54, 13, 7, 12, 14, 0, 6, 25, 43. Zjistěte horní odhad směrodatné odchylky rozdělení životnosti za předpokladu normálního rozdělení na hladině spolehlivosti 90 %. Protože rozdělení životnosti pravděpodobně normální není, zkuste generovat posloupnosti 10 čísel, sestávající se z výše naměřených hodnot, kde každá hodnota je stejně pravděpodobná. U každé posloupnosti určete odhad S směrodatné odchylky a horní odhad určete z kvantilu empirické distribuční funkce. *Tento postup se nazývá bootstrap.*

HY7) Vygenerujte 50 čísel s Poissonovým rozdělením (např. funkce POIDEV z Numerical Recipes) s parametrem $\lambda = 1.75$. Z těchto dat určete parametr Poissonova rozdělení a pomocí χ^2 testu ověřte, že se jedná o Poissonovo rozdělení.

FT1) Signál tvaru $f(x) = \cos[(\omega_0 + a\cos\omega t)t]$ s $\omega = 5,\ \omega_0 = 0.7,\ a = 0.05$ vypočtěte v nejméně 500 bodech v intervalu cca (0,10). Pomocí rychlé Fourierovy transformace určete autokorelační funkci a energetické spektrum. Prozkoumejte, jak se mění spektrum při změně délky intervalu. Vyzkoušejte rozdělení do podintervalů a různá okna.

FT2) Proveďte vyhlazení křivky pomocí rychlé Fourierovy transformace. Je zadáno 32 hodnot x od 0 do 15.5 v intervalu 0.5. Hodnoty y jsou zadány v tabulce

1.15	2.40	2.34	2.40	3.54	0.835	-0.057	-0.369
-0.375	-0.27	0.68	1.95	0.887	4.82	2.34	2.52
0.85	-0.356	-2.36	-1.43	-2.36	-2.08	0.015	1.85
6.35	3.95	6.40	6.11	2.43	-0.01	-2.44	-8.73

Vyzkoušejte pro několik hodnot max. frekvence, nad kterou jsou vyšší potlačovány. Porovnejte výsledek při ostré a "rozmazané" hranici potlačovaných frekvencí.

FT3) Proveďte dekonvoluci pomocí FFT (rychlé Fourierovy transformace). Použijte normálního filtrování k odstranění šumu. Přístrojová funkce je v časech t = 0,1,...,8 dána hodnotami v tabulce

0	0.076	0.141	0.183	0.2	0.183	0.141	0.076	0

Signál, naměřený pro časy t = 0, 1, ..., 23 je zapsán v následující tabulce

0.000	0.225	0.621	0.647	0.598	0.838	1.360	1.046
1.458	1.597	1.696	2.134	1.753	1.030	1.747	0.979
0.947	0.802	0.315	0.706	0.288	0.193	0.095	0.084

FT4) Vstupní signál je $f(t) = 0$ pro $t < 0$ a $f(t) = \exp(-t)$ pro $t \geq 0$. Přístrojová funkce je

$$g(t) = \begin{cases} 0 & t < 0 \wedge t > 1 \\ 16t^2 & t \in \langle 0, 0.25 \rangle \\ 2 - 16(t-0.5)^2 & t \in \langle 0.5, 0.75 \rangle \\ 16(t-1)^2 & t \in \langle 0.75, 1 \rangle \end{cases}$$

Na výstupu je signál konvolucí $f * g$ s náhodným šumem. Předpokládejte, že zašuměný signál je

$$\tilde{h}(t) = f * g + 0.1 * a(t) * \sqrt{f * g}$$

kde $a(t)$ je náhodné číslo s rovnoměrným rozdělením v intervalu $\langle -1, 1 \rangle$.
Předpokládejte interval vzorkování $\Delta t = 0.02$ a nagenerujte zašuměný signál v cca 200 bodech. Proveďte dekonvoluci s odfiltrováním šumu.

FT5) Na signál $y(t) = \sin(5t+1) + \cos 15t + \sin 30t$ vzorkovaný s $\Delta t = 0.003$ použijte jednoduchý rekurzivní filtr - pásmovou propust pro frekvence 10 - 20 Hz sestrojený dle prednášky. Spočtěte vyfiltrovaný signál a jeho spektrum.

FT6) Nagenerujte periodický signál se šumem $f(t) = \sin(5t) + e(t)$, kde $e(t)$ je náhodné číslo s rozdělením N(0,2) (např. funkce GASDEV z Numerical Recipes). Pomocí autokorelace ze signálu určete amplitudu a frekvenci periodické funkce.

NP1) Použijte generátor náhodných čísel a generujte posloupnost $y(t) = \sin \omega t + \xi(t)$ pro časy $t = k \cdot \Delta t$, kde $\xi(k \cdot \Delta t)$ jsou vzájemně nezávislá náhodná čísla se stejnoměrným rozdělením v intervalu (-1,1). Spočtěte autokorelační funkci, zkuste oddělit signál od šumu na základu předpokladu o periodickém signálu.

NP2) Použijte procedury GASDEV pro výpočet náhodných čísel s normálním rozdělením k simulaci náhodného procesu. Nechť je počáteční hodnota x_0 daná normálním rozdělením s nulovou střední hodnotou a jednotkovým rozptylem. Nechť je hodnota x_{k+1} dána normálním rozdělením s jednotkovým rozptylem a se střední hodnotou $\mu_{k+1} = x_k \ \exp(-\tau/\tau_0)$, kde x_k je hodnota signálu v čase $k\tau$. Pro vygenerovaný signál vypočtěte autokorelační funkci a energetické spektrum. Zvolte vhodný krok τ.

NP3) Pomocí generátoru náhodných čísel simulujte náhodný proces spočívající v házení 2 spravedlivými kostkami, kde hodnota v čase k je součtem bodů na obou kostkách a hodnoty na kostce s větší hodnotou v předchozím čase. Najděte autokorelační funkci a výkonové spektrum.

NP4) Pomocí generátoru náhodných čísel simulujte zobecněný telegrafní signál. Spočtete autokorelační funkci signálu a jeho energetické spektrum. Proveďte středování přes několik náhodně vygenerovaných signálů.

NP5) Pomocí generátoru náhodný čísel simulujte homogenní Markovův řetězec s 5 stavy (-2, -1, 0, 1, 2) a s následujícími pravděpodobnostmi přechodů za čas Δt (označme $\lambda = p\Delta t$) - pravděpodobnost přechodu do sousedního stavu je $0.5 \times (1 - \exp(-\lambda))$, pravděpodobnost zůstat v daném stavu je zbytek do 1 (stavy -1, 0, 1 mají 2 sousední stavy, stavy -2, 2 mají jen jeden sousední stav). V čase t=0 je pravděpodobnost všech stavů 0.2.
Ukažte, že se jedná o centrovaný stacionární náhodný proces. Spočítejte autokorelační funkci signálu a dobu korelace. Vyzkoušejte pro několik hodnot Δt.

NP6) Pomocí generátoru náhodných čísel simulujte zobecněný telegrafní signál se zpožděním. Jsou dva stavy +1, -1 s pravděpodobností p=0.5 v čase t=0, pravděpodobnost přechodu mezi nimi za jednotku času q= 0.5, po každém přechodu po dobu $\Delta = 0.5$ je pravděpodobnost přechodu nulová. Spočtete autokorelační funkci signálu a jeho energetické spektrum. Proveďte středování přes několik náhodně vygenerovaných signálů.

NP7) Pomocí generátoru náhodných čísel simulujte počet nerozpadlých jader pro proces rozpadu s rychlostí rozpadu q = 2 s^{-1}, jestliže na počátku je $N_0 = 50$ nebo $N_0 = 500$ jader. Vygenerujte několik realizací náhodného procesu a zkoumejte časovou závislost a autokorelační funkci odchylky náhodných hodnot od teoretické střední závislosti $N(t) = N_0\ \exp(-qt)$.

NP8) Pomocí generátoru náhodných čísel simulujte náhodný proces - počet lidí stojících ve frontě na obsluhu. Pravděpodobnost příchodu člověka je q = 0.05 s^{-1}, obsluha obslouží člověka za $\tau = 18s$. Na začátku není ve frontě nikdo, po určité době proces konverguje ke stacionárnímu. Určete rozdělení pravděpodobnosti počtu čekajících, autokorelační funkci a podíl času, kdy je obsluha bez práce.

NP9) Pomocí generátoru náhodných čísel s normálním rozdělěním (např. funkce GASDEV z Numerical Recipes) simulujte Markovovský náhodný proces takový, že v čase t=0 je hustota pravděpobnosti $f(y|t=0) = \exp(-y^2/2)/\sqrt{2\pi}$ - tj. rozdělení $N(0,1)$. Pokud byla v čase t hodnota náhodného procesu y_1, pak podmíněná hustota pravděpodobnosti hodnoty y_2 v čase $t+\tau$ je

$$f(y_2|y_1) = \frac{1}{\sqrt{2\pi[1-\exp(-2\tau)]}}\ \exp\left\{-\frac{[y^2 - y_1\exp(-\tau)]^2}{2[1-\exp(-2\tau)]}\right\}$$

Zvolte $\tau = 0.05$. Určete střední hodnotu, disperzi a autokorelační funkci vygenerovaného signálu.

EX1) (pro LTO) Nastudujte možnosti regresní analýzy v tabulkovém kalkulátoru EXCEL. Ukažte na příkladech.

EX2) (pro LTO) Nastudujte možnosti provádění Studentových t-testů a analýzy korelace v tabulkovém kalkulátoru EXCEL. Ukažte na příkladech.

EX3) (pro LTO) Nastudujte možnosti analýzy a zpracování časových řad v tabulkovém kalkulátoru EXCEL. Ukažte na příkladech.

GR1) Seznamte se s některými metodami průzkumové grafické analýzy dat (např. stonek s listy, krabicový graf, vrubové grafy, hvězdicové a paprskové diagramy) a ukažte je na příkladech.