

Regresní analýza

1. Byla zjištěna výška otců a výška jejich nejstarších synů [v cm].

otec	165	178	158	170	180	160	170	167	185	165	173	175
syn	162	184	163	170	189	165	177	170	187	176	171	183

- Sestrojte bodový graf.
- Určete regresní přímku a nakreslete její graf.
- Odhadněte průměrnou výšku syna při výšce otce 178 cm.
- Určete reziduální rozptyl, směrodatné chyby těchto odhadů regresních parametrů a zkonstruuje 95% intervaly spolehlivosti pro parametry regresní přímky.
- Odhadněte výšku syna při výšce otce 180 cm, určete interval 95% interval spolehlivosti pro regresní funkci i pro predikci v tomto bodě.
- Testujte významnost regresních parametrů na hladině významnosti 0,05.
- Proveďte test významnosti regresního modelu, určete koeficient determinace.

[Datový soubor: vyska_otec_syn.txt]

2. O 7 vybraných strojích v určitém podniku máme informace o jejich stáří (v letech) a týdenních nákladech na jejich údržbu (v Kč):

stáří stroje	1	1	3	3	5	6	7
náklady	35	52	81	105	100	125	120

- Sestrojte bodový graf.
- Určete regresní přímku a nakreslete její graf.
- Určete regresní logaritmickou křivku a nakreslete její graf.
- Určete bodové odhady parametrů regresní přímky, reziduální rozptyl, směrodatné chyby těchto odhadů a zkonstruuje 95% intervaly spolehlivosti pro parametry regresní přímky.
- Určete bodové odhady regresní logaritmické křivky, reziduální rozptyl, směrodatné chyby těchto odhadů a zkonstruuje 95% intervaly spolehlivosti pro parametry regresní přímky.
- Pro oba modely určete interval 95% interval spolehlivosti pro regresní funkci i pro predikci pro stáří stroje 4 roky.
- Testujte významnost parametrů regresní přímky a regresní logaritmické křivky na hladině významnosti 0,05.
- Pro oba modely proveďte test významnosti regresního modelu, určete koeficienty determinace. Rozhodněte, který z uvedených modelů je pro popis studované závislosti vhodnější, zdůvodněte.

[Datový soubor: stari_stroje_naklady.txt]

3. Zajímáme se o brzdovou dráhu 63 automobilů v závislosti na výchozí rychlosti. K dispozici je celkem $n = 63$ měření. Proměnná rychlost udává výchozí rychlost (míle/hod.) před začátkem brzdění, proměnná dráha pak udává odpovídající brzdovou dráhu uvedenou ve stopách.

Rychlost	4	5	5	5	5	7	7	8	8	8
Dráha	4	2	8	8	4	6	7	9	8	13
Rychlost	8	9	9	9	10	10	10	12	12	12
Dráha	11	5	13	5	8	17	14	11	21	19
Rychlost	13	13	13	14	14	15	16	16	16	17
Dráha	18	27	15	14	16	16	19	14	34	29
Rychlost	17	18	18	18	19	20	21	21	21	22
Dráha	22	47	29	34	30	48	55	39	42	35
Rychlost	24	25	25	25	25	26	26	27	27	28
Dráha	56	33	59	48	56	39	41	78	57	64
Rychlost	28	29	29	30	30	30	31	35	35	36
Dráha	84	68	54	60	101	67	77	85	107	79
Rychlost	39	40	40							
Dráha	138	110	134							

- Sestrojte bodový graf.
- Určete regresní přímku a nakreslete její graf.
- Určete regresní parabolu a nakreslete její graf.
- Odhadněte brzdnu dráhu pro rychlost 25 mil/hod.
- Určete bodové odhady parametrů regresní přímky, reziduální rozptyl, směrodatné chyby těchto odhadů a zkonstruujte 95% intervaly spolehlivosti pro parametry regresní přímky.
- Určete bodové odhady parametrů kvadratické regresní funkce, reziduální rozptyl, směrodatné chyby těchto odhadů a zkonstruujte 95% intervaly spolehlivosti pro parametry této regresní funkce.
- Odhadněte brzdnu dráhu pro rychlost 20 mil/hod pomocí obou modelů. Pro oba modely určete interval 95% interval spolehlivosti pro regresní funkci i pro predikci pro rychlost 20 mil/hod.
- Testujte významnost těchto parametrů obou modelů na hladině významnosti 0,05.
- Pro oba modely proveďte test významnosti regresního modelu, určete koeficienty determinace. Rozhodněte, který z uvedených modelů je pro popis studované závislosti vhodnější, zdůvodněte.

[Datový soubor: brzdna_draha.txt]

- V padesátých letech došlo k úniku radioaktivního odpadu ze skládky v Hanfordu ve státě Washington do řeky Columbia River. V devíti okrcích níže po proudu ve státě Oregon bylo počítáno vystavení radioaktivitě X (na základě vzdálenosti od Hanfordu a vzdálenosti průměrného obyvatele od řeky apod.). Současně se sledovala úmrtnost na rakovinu Y (úmrtnost na 100 000 lidí za rok v letech 1959–64). Získané údaje jsou shrnuty v následující tabulce. Pro daný datový soubor odhadněte parametry těchto modelů:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \dots, n$$

$$Y_i = \beta_0 + \beta_1 \ln x_i + \epsilon_i, i = 1, \dots, n.$$

okrsek	radioaktivní vystavení X	úmrtnost na rakovinu Y
Clatsop	8,3	210
Columbia	6,4	180
Cilliam	3,4	130
Hood River	3,8	170
Morrow	2,6	130
Portland	11,6	210
Sherman	1,2	120
Umatilla	2,5	150
Wasco	1,6	140

- Pro regresní přímku určete reziduální rozptyl, směrodatné chyby odhadů regresních parametrů a zkonstruuje 95% intervaly spolehlivosti pro tyto parametry.
- Pro logaritmickou regresní křivku určete reziduální rozptyl, směrodatné chyby odhadů regresních parametrů a zkonstruuje 95% intervaly spolehlivosti pro tyto parametry.
- Odhadněte úmrtnost pro hodnotu radioaktivní vystavení 8 pomocí obou modelů. Pro oba modely určete interval 95% interval spolehlivosti pro regresní funkci i pro predikci pro hodnotu radioaktivní vystavení 8.
- Pro oba modely zkonstruuje pásy spolehlivosti pro regresní funkce a predikované hodnoty, výsledky zobrazte graficky.
- Testujte významnost parametrů těchto modelů na hladině významnosti 0,05.
- Pro oba modely proveďte test významnosti regresního modelu, určete koeficienty determinace. Rozhodněte, který z uvedených modelů je pro popis studované závislosti vhodnější, zdůvodněte.

[Datový soubor: radiace_umrtnost.txt]

- Cílem studie bylo nalézt závislost mezi tělesným tukem lehkých atletů-běžců y , kteří trénují asi 12 hodin, a zkonsumovaným tukem v jejich každodenní stravě x . U náhodného vzorku 18 běžců byl měřen jejich tělesný podkožní tuk y [%] a sledován v závislosti na zkonsumovaném tuku ve stravě x [%].

x	22,0	30,0	24,0	22,0	21,0	36,0	14,0	17,0	20,0
y	9,80	9,70	12,00	11,70	11,60	11,60	8,00	8,60	10,40
x	21,0	35,0	37,0	32,0	35,0	35,0	26,0	24,0	14,0
y	9,70	11,20	10,80	10,90	12,30	11,50	7,80	10,20	7,90

Pro daný datový soubor odhadněte parametry těchto modelů:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \dots, n$$

$$Y_i = \beta_0 + \frac{\beta_1}{x_i} + \epsilon_i, i = 1, \dots, n.$$

- Pro regresní přímku určete reziduální rozptyl, směrodatné chyby odhadů regresních parametrů a zkonstruuje 95% intervaly spolehlivosti pro tyto parametry.
- Pro hyperbolickou regresní křivku určete reziduální rozptyl, směrodatné chyby odhadů regresních parametrů a zkonstruuje 95% intervaly spolehlivosti pro tyto parametry.
- Odhadněte množství podkožního tuku pro hodnotu tuku ve stravě 30 % pomocí obou modelů. Pro oba modely určete interval 95% interval spolehlivosti pro regresní funkci i pro predikci hodnotu tuku ve stravě 30 %.

- d) Pro oba modely zkonstruujte pásy spolehlivosti pro regresní funkce a predikované hodnoty, výsledky zobrazte graficky.
- e) Pro regresní přímku testujte významnost regresních parametrů na hladině významnosti 0,05 a 0,01.
- f) Pro hyperbolickou regresní křivku testujte významnost regresních parametrů na hladině významnosti 0,05 a 0,01.
- g) Pro oba modely proveďte test významnosti regresního modelu, určete koeficienty determinace. Rozhodněte, který z uvedených modelů je pro popis studované závislosti vhodnější, zdůvodněte.

[Datový soubor: tuk_sportovci.txt]

6. U automobilu byla změřena spotřeba Y v závislosti na rychlosti X . Údaje jsou v tabulce

Rychlost [km/hod.]	40	50	60	70	80	90	100	110
Spotřeba [l/100 km]	5,7	5,4	5,2	5,2	5,8	6,0	7,5	8,1

Pro daný datový soubor odhadněte parametry těchto modelů:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \dots, n$$

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i, i = 1, \dots, n.$$

- a) Pro regresní přímku určete reziduální rozptyl, směrodatné chyby odhadů regresních parametrů a zkonstruujte 95% intervaly spolehlivosti pro tyto parametry.
- b) Pro parabolickou regresní křivku určete reziduální rozptyl, směrodatné chyby odhadů regresních parametrů a zkonstruujte 95% intervaly spolehlivosti pro tyto parametry.
- c) Odhadněte spotřebu pro hodnotu rychlosti 75 km/hod. pomocí obou modelů. Pro oba modely určete interval 95% interval spolehlivosti pro regresní funkci i pro predikci pro hodnotu rychlosti 75 km/hod.
- d) Pro oba modely zkonstruujte pásy spolehlivosti pro regresní funkce a predikované hodnoty, výsledky zobrazte graficky.
- e) Pro regresní přímku testujte významnost regresních parametrů na hladině významnosti 0,05 a 0,01.
- f) Pro parabolickou regresní křivku testujte významnost regresních parametrů na hladině významnosti 0,05 a 0,01.
- g) Pro oba modely proveďte test významnosti regresního modelu, určete koeficienty determinace. Rozhodněte, který z uvedených modelů je pro popis studované závislosti vhodnější, zdůvodněte.

[Datový soubor: rychlost_spotreba.txt]

7. Data popisují výsledky vstupních zdravotních testů uchazečů o službu u policie.

Tlak	66	87	85	59	76	77	70	66	75	66
Hmotnost	87,36	117,6	82,85	62,32	82	102	70,12	88,07	77,96	74,33
Tuk	16,98	27,6	6,61	3,26	19	27	6,88	18,8	18,87	8,15
Tlak	74	68	72	76	94	63	80	67	77	78
Hmotnost	56,2	81,75	80,24	74,81	61,98	95,23	72,48	92,45	104,56	66,2
Tuk	3,44	20,31	12,96	12,42	3,58	12,91	11,34	17,5	18,93	10,94
Tlak	77	67	78	78	80	95	76	78	73	80
Hmotnost	87,16	82,42	64,11	81,57	99,85	78,49	87,13	65,64	51,76	67,14
Tuk	17,72	9,55	9,54	13,1	17,75	9,57	18,52	6,4	2,86	4,31

Tlak	81	61	65	69	66	75	72	66	93	77
Hmotnost	78,74	86,83	70,48	72,67	85,86	84,86	66,97	68,33	63,34	85,72
Tuk	16,26	9,72	6,29	4,37	14,43	17	5,8	8,14	3,63	23,61
Tlak	68	71	84	81	74	79	89	79	80	67
Hmotnost	89	95,17	84,19	63,12	70,01	82,11	71	94,56	70,91	79,19
Tuk	18,83	19,16	15,83	8,77	6,61	22,22	8,29	26,82	9,32	19,9

Popište vhodným regresním modelem (pokud to lze) závislost tlaku na hmotnosti a procentech tuku v těle. Najděte vhodný model pro popis závislosti hmotnosti na procentech tuku v těle.

- Určete reziduální rozptyl, směrodatné chyby odhadů regresních parametrů a zkonstruujte 95% intervaly spolehlivosti pro oba modely.
- Pro model popisující závislost hmotnosti na procentu tuku v těle odhadně hmotnost pro hodnotu tuku 20 %. Spočtete 95% intervaly spolehlivosti pro hodnotu regresní funkce i predikovanou hodnotu v tomto bodě.
- Testujte významnost regresních parametrů pro oba modely na hladině významnosti 0,05.
- Pro oba modely proveďte test významnosti regresního modelu, určete koeficienty determinace. Rozhodněte, který z uvedených modelů je pro popis studované závislosti vhodný, případně který nikoli. Svě závěry zdůvodněte.

[Datový soubor: vstupni_testy.txt]