

UNIVERZITA PARDUBICE

**Fakulta chemicko-technologická
Katedra analytické chemie**

Licenční studium chemometrie na téma

Statistické zpracování dat

Semestrální práce ze 6. soustředění

Předmět: 3.3 Tvorba nelineárních regresních modelů v analýze dat

Vedoucí licenčního studia: Prof. RNDr. Milan Meloun, DrSc.

Vypracoval: Ing. Radek Novotný

Licenční studium Statistické zpracování experimentálních dat.

Předmět: 3.3 Tvorba nelineárních regresních modelů v analýze dat

Přednášející: Prof. RNDr. Milan Meloun, DrSc., Prof. Ing. Jiří Militký, CSc.

Zadání: Vypracovat alespoň dvě úlohy, nejlépe z dat vlastního pracoviště.

Úloha 1.

Porovnání růstových modelů obsažených v programu ADSTAT a nalezení nejvhodnějšího modelu pro zadaná data.

Zadání: Pro data ze smrkového porostu z Orlických hor vybrat nejvhodnější růstový model obsažený v modulu „Růstové křivky“ programu ADSTAT. Závisle proměnná veličina je výška stromů, nezávisle proměnná věk. Vstupní data jsou v souboru 3_3-uloha1.xls.

Řešení: Úloha byla řešena v programu ADSTAT. Kvůli potížím s kopírováním grafů byla úloha vyřešena paralelně také pomocí SW QCExpert. Jako vstupní odhady parametrů v QCE byly použity hodnoty nalezené v ADSTAT. Uvedeny jsou všechny výsledky, ve výstupech z QCE jsou na obr. 1 až 4 vloženy grafy.

Modul „Růstové křivky“ obsahuje následující růstové modely:

- Schnuteho model

$$y = p1 \cdot \left(1 + p4 \cdot e^{-p3 \cdot e^{-p2 \cdot x}} \right)^{\frac{1}{p4}}; \text{ pro } p4 > 1$$

- Mitscherlichův model

$$y = p1 \cdot \left(1 + p4 \cdot e^{-p3 \cdot e^{-p2 \cdot x}} \right)^{\frac{1}{p4}}; \text{ pro } p4 = -1$$

- Richardsův model

$$y = p1 \cdot \left(1 + p4 \cdot e^{-p3 \cdot e^{-p2 \cdot x}} \right)^{\frac{1}{p4}}; \text{ pro } p4 < 1$$

- Gompertzův model

$$y = p1 \cdot e^{-e^{-p2 \cdot e^{-p3 \cdot x}}}$$

- Logistický model

$$y = p1 \cdot \left(1 + p4 \cdot e^{-p3 \cdot e^{-p2 \cdot x}} \right)^{\frac{1}{p4}}; \text{ pro } p4 = 1$$

ADSTAT

Odhadování parametrů

Schnuteho model:

| Parametr | Bodový odhad | Směrodatná odchylka | Absolutní vychýlení | Relativní vychýlení[%] |
|----------|--------------|---------------------|---------------------|------------------------|
| p[1] | 1.9142E+03 | 1.0806E+02 | 1.1735E+01 | 6.1304E-01 |
| p[2] | 3.0748E+01 | 3.1077E-01 | 2.2570E-02 | 7.3401E-02 |
| p[3] | 7.5498E-02 | 9.7603E-03 | 1.9472E-04 | 2.5791E-01 |
| p[4] | 3.1919E-01 | 1.5699E-01 | 2.8213E-03 | 8.8391E-01 |

Mitscherlichův model:

| Parametr | Bodový odhad | Směrodatná odchylka | Absolutní vychýlení | Relativní vychýlení[%] |
|----------|--------------|---------------------|---------------------|------------------------|
| p[1] | 3.3942E+05 | 7.4154E+07 | 1.6181E+10 | 4.7672E+06 |
| p[2] | 3.0379E+01 | 2.1383E+00 | 1.8790E-03 | 6.1852E-03 |
| p[3] | 2.2471E-04 | 4.9083E-02 | 1.0730E-02 | 4.7751E+03 |

Richardsův model:

| Parametr | Bodový odhad | Směrodatná odchylka | Absolutní vychýlení | Relativní vychýlení[%] |
|----------|--------------|---------------------|---------------------|------------------------|
| p[1] | 1.9142E+03 | 1.0806E+02 | 1.1735E+01 | 6.1304E-01 |
| p[2] | 3.0748E+01 | 3.1077E-01 | 2.2570E-02 | 7.3401E-02 |
| p[3] | 7.5498E-02 | 9.7603E-03 | 1.9472E-04 | 2.5791E-01 |
| p[4] | 3.1919E-01 | 1.5699E-01 | 2.8213E-03 | 8.8391E-01 |

Gompertzův model:

| Parametr | Bodový odhad | Směrodatná odchylka | Absolutní vychýlení | Relativní vychýlení[%] |
|----------|--------------|---------------------|---------------------|------------------------|
| p[1] | 2.2009E+03 | 5.0213E+01 | 1.6681E+00 | 7.5792E-02 |
| p[2] | 5.5754E-02 | 1.3052E-03 | 2.7464E-06 | 4.9260E-03 |
| p[3] | 3.0887E+01 | 4.4458E-01 | 1.3642E-02 | 4.4168E-02 |

Logistický model:

| Parametr | Bodový odhad | Směrodatná odchylka | Absolutní vychýlení | Relativní vychýlení[%] |
|----------|--------------|---------------------|---------------------|------------------------|
| p[1] | 1.6261E+03 | 2.1361E+01 | 5.4802E-01 | 3.3701E-02 |
| p[2] | 3.0962E+01 | 2.6434E-01 | 6.0490E-03 | 1.9537E-02 |
| p[3] | 1.1737E-01 | 1.8545E-03 | 8.0635E-06 | 6.8700E-03 |

Základní statistické charakteristiky

| model | RSC | Regresní rabat |
|-------------------|-------------------|-------------------|
| Schnuteho | 1.8847E+03 | 9.9942E+01 |
| Mitscherlichův | 8.3255E+06 | -1.5532E+02 |
| Richardsův | 1.1273E+03 | 9.9965E+01 |
| Gompertzův | 1.3273E+03 | 9.9959E+01 |
| Logistický | 1.8847E+03 | 9.9942E+01 |

QCExpert

Odhadování parametrů

Schnuteho model:

| Odhady parametrů | Parametr | Směr. odchylka | Dolní mez | Horní mez |
|------------------|----------|----------------|-----------|-----------|
| P1 | 1914.301 | 108.0025 | 1691.395 | 2137.207 |
| P2 | 30.74985 | 0.310664 | 30.10867 | 31.39103 |
| P3 | 0.075494 | 0.009753 | 0.055363 | 0.095624 |
| P4 | 0.319265 | 0.156892 | -0.00454 | 0.643075 |

Mitscherlichův model:

| Odhady parametrů | Parametr | Směr. odchylka | Dolní mez | Horní mez |
|------------------|----------|----------------|-----------|-----------|
| P1 | 339417.2 | 77562206 | -1.6E+08 | 1.6E+08 |
| P2 | 578.504 | 114903.1 | -236570 | 237726.9 |
| P3 | 0.000239 | 0.050538 | -0.10407 | 0.104545 |
| P4 | -0.87264 | 1.088077 | -3.11832 | 1.373043 |

Richardsův model:

| Odhady parametrů | Parametr | Směr. odchylka | Dolní mez | Horní mez |
|------------------|----------|----------------|-----------|-----------|
| P1 | 1914.301 | 108.0025 | 1691.395 | 2137.207 |
| P2 | 30.74985 | 0.310664 | 30.10867 | 31.39103 |
| P3 | 0.075494 | 0.009753 | 0.055363 | 0.095624 |
| P4 | 0.319265 | 0.156892 | -0.00454 | 0.643075 |

Gompertzův model:

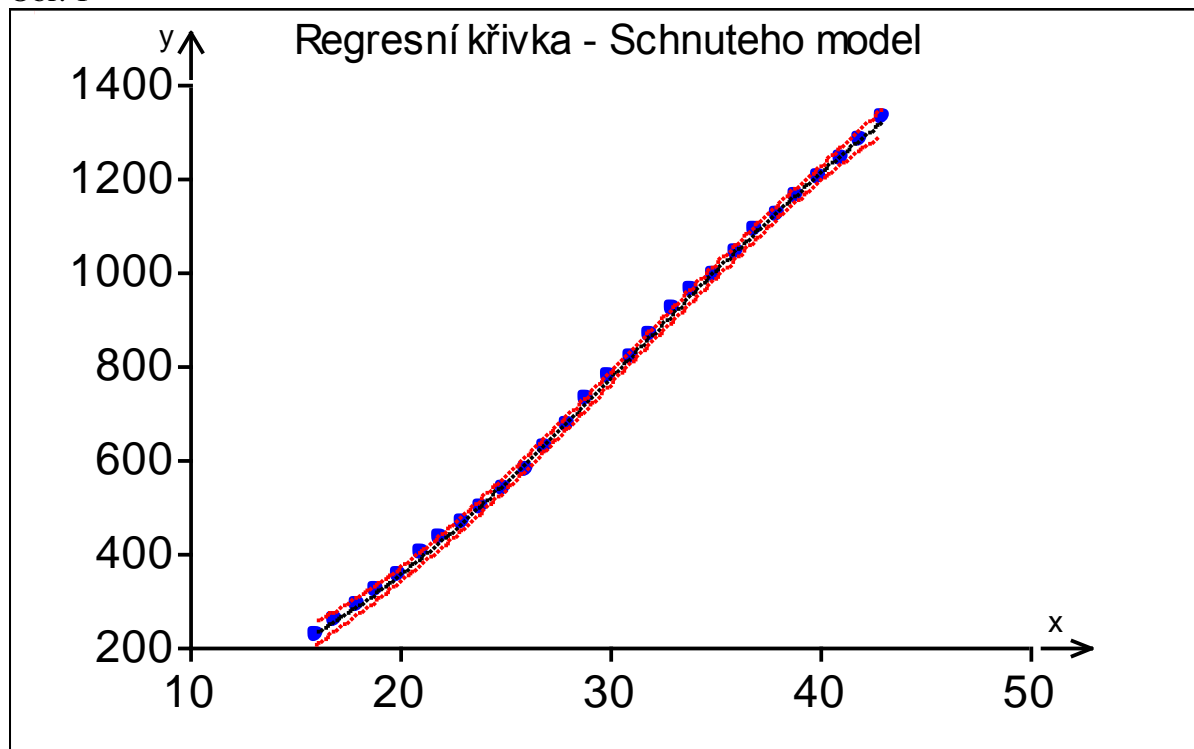
| Odhady parametrů | Parametr | Směr. odchylka | Dolní mez | Horní mez |
|------------------|----------|----------------|-----------|-----------|
| P1 | 2201.156 | 50.19433 | 2097.779 | 2304.534 |
| P2 | 0.055747 | 0.001305 | 0.05306 | 0.058434 |
| P3 | 30.88944 | 0.444489 | 29.974 | 31.80489 |

Logistický model:

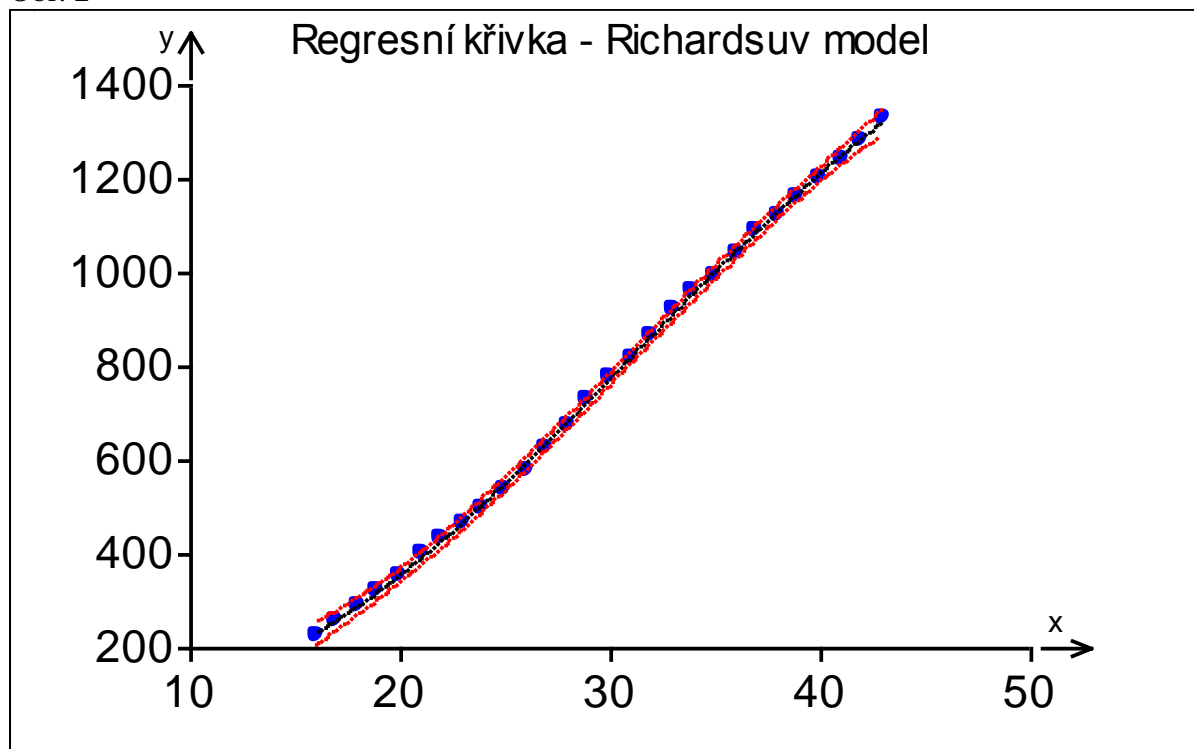
| Odhady parametrů | Parametr | Směr. odchylka | Dolní mez | Horní mez |
|------------------|----------|----------------|-----------|-----------|
| P1 | 1914.301 | 108.0025 | 1691.395 | 2137.207 |
| P2 | 30.74985 | 0.310664 | 30.10867 | 31.39103 |
| P3 | 0.075494 | 0.009753 | 0.055363 | 0.095624 |
| P4 | 0.319265 | 0.156892 | -0.00454 | 0.643075 |

Grafy regresních křivek

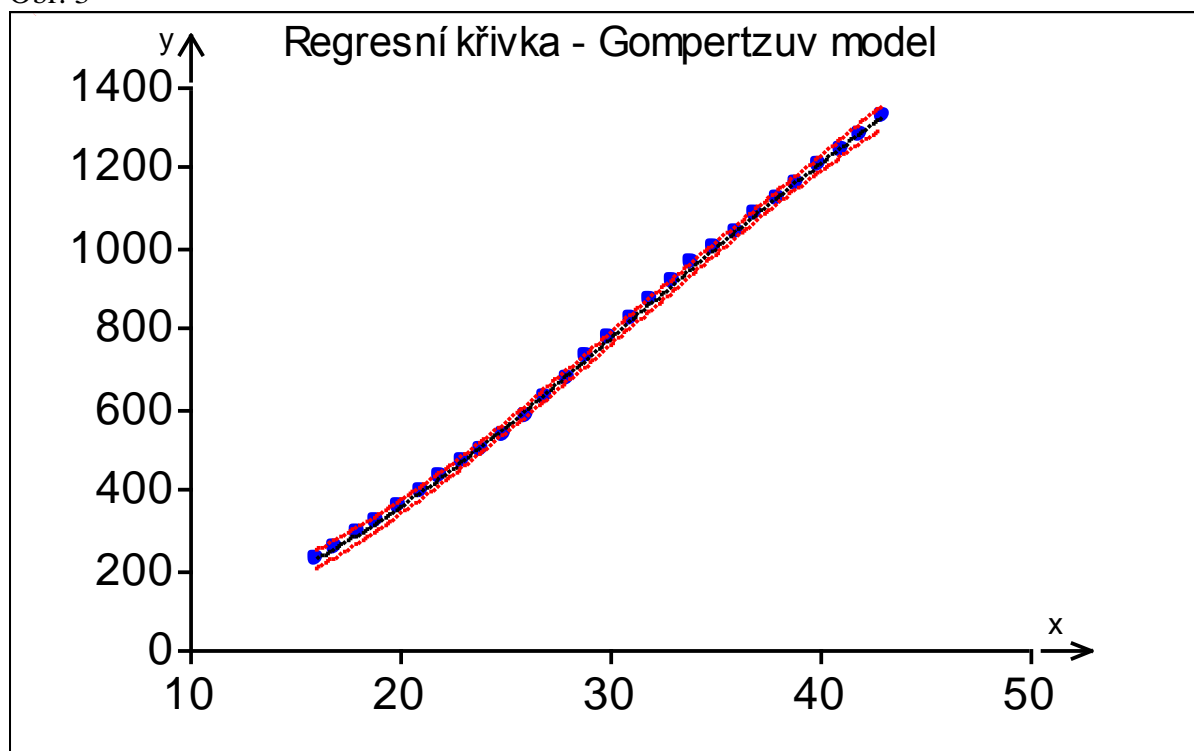
Obr. 1



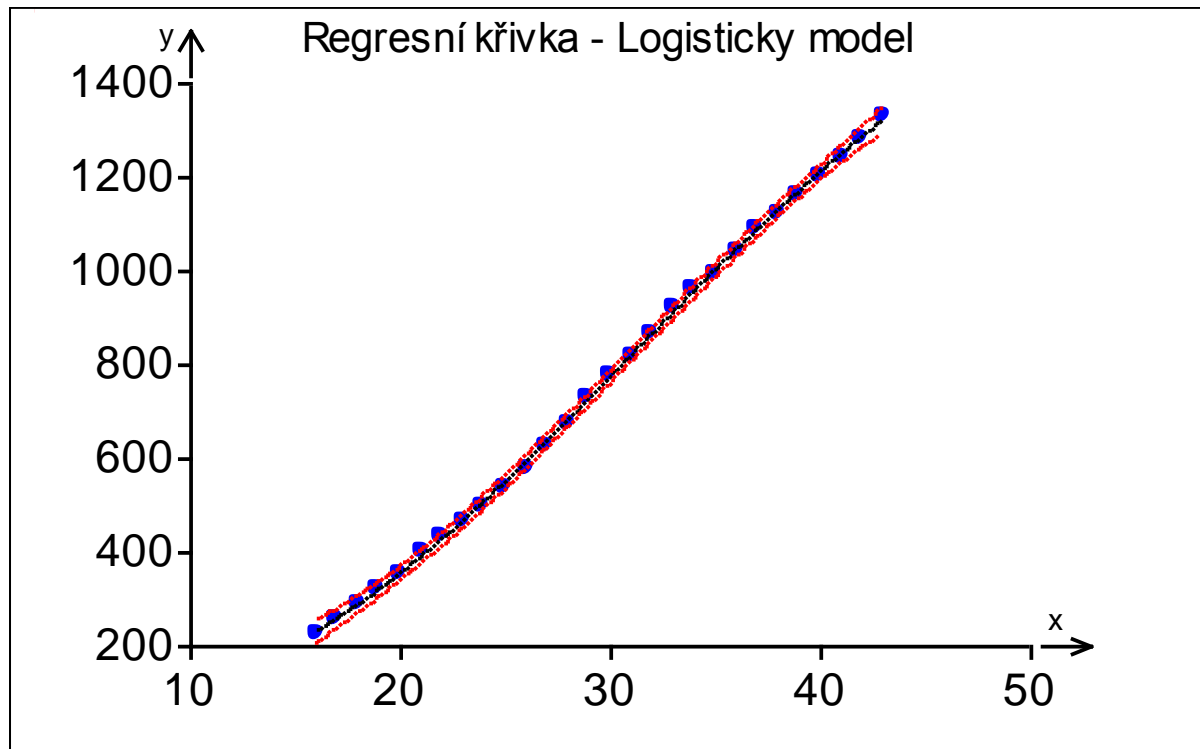
Obr. 2



Obr. 3



Obr. 4



Základní statistické charakteristiky

| | Schnute | Mitscherlich | Richards | Gompertz | Logistický |
|----------------------------------|--------------------|--------------|--------------------|-------------|--------------------|
| Vícenásobný korel. koef. R: | 0.999827286 | 0.999330013 | 0.999827286 | 0.999796632 | 0.999827286 |
| Koeficient determinace R^2 : | 0.999654601 | 0.998660475 | 0.999654601 | 0.999593305 | 0.999654601 |
| Predikovaný korel. koef. R_p : | 0.999489304 | 0.998065567 | 0.999489304 | 0.999508633 | 0.999489304 |
| Stř. kvadr. chyba predikce MEP: | 59.47184015 | 225.2696228 | 59.47184017 | 57.22095296 | 59.47184009 |
| Akaikeho informační kritérium: | 111.4439762 | 149.3943568 | 111.4439762 | 114.0181843 | 111.4439762 |

Podle statistických charakteristik prokládají naměřenými daty křivku shodně tři modely.

Závěr:

Jako nejlepší pro data naměřená ve smrkovém porostu v Orlických horách se jeví Richardsův růstový model.

Úloha 2.

Nalezení parametrů Korfovy růstové funkce pro data z mladého smrkového porostu z Orlických hor

Zadání: V roce 2002 byly v Orlických horách měřeny výškové přírůsty v mladých smrkových porostech. Pro úlohu byla vybrána data z porostu 360 C4. Vstupní data jsou v souboru 3_3-uloha1.xls. Úkolem je nalézt nejlepší odhady parametrů Korfovy růstové funkce pro tato data.

Řešení: Úloha byla řešena v programu ADSTAT.

Korfova růstová funkce:

$$y = A \cdot e^{\frac{b}{(-n) \cdot t^{n-1}}}$$

y ... závisle proměnná (výška)

A, b, n ... parametry funkce

t ... věk stromu

A. Exploratorní analýza dat**(1) KLASICKÉ ODHADY PARAMETRŮ**

| | | | |
|-------------------|------------|-------------------|------------|
| Medián: | 7.5040E+02 | Průměr: | 7.5548E+02 |
| Rozptyl: | 1.2077E+05 | III.cent. moment: | 2.5318E+06 |
| IV.cent. moment : | 2.3869E+10 | Šikmost: | 6.1431E-02 |
| Špičatost: | 1.6971E+00 | Směrodatná odch.: | 3.4752E+02 |

(2) KVANTILY A PÍSMENOVÉ HODNOTY:

Kvantilové míry:

| Procento | Kvantil | Procento | Kvantil |
|----------|------------|----------|------------|
| 5 | 2.6504E+02 | 10 | 3.1029E+02 |
| 15 | 3.5553E+02 | 20 | 4.1118E+02 |
| 25 | 4.5885E+02 | 30 | 5.0036E+02 |
| 35 | 5.5310E+02 | 40 | 6.1712E+02 |
| 45 | 6.8081E+02 | 50 | 7.5040E+02 |
| 55 | 8.1183E+02 | 60 | 8.7696E+02 |
| 65 | 9.4004E+02 | 70 | 9.9219E+02 |
| 75 | 1.0495E+03 | 80 | 1.1072E+03 |
| 85 | 1.1583E+03 | 90 | 1.2128E+03 |
| 95 | 1.2666E+03 | | |

Písmenové hodnoty:

| Kvantil | Písmeno | Pravděpodobnost | Spodní mez | Horní mez |
|---------|---------|-----------------|------------|------------|
| Sedecil | D | 0.0625 | 2.7665E+02 | 1.2518E+03 |
| Oktil | E | 0.1250 | 3.3259E+02 | 1.1863E+03 |
| Kvartil | F | 0.2500 | 4.5885E+02 | 1.0495E+03 |
| Medián | M | 0.5000 | 7.5040E+02 | 7.5040E+02 |

(3) KVANTILOVÉ MÍRY:

| | | | |
|--------------|------------|------------|------------|
| Kvantil: | F (0.25) | E (0.125) | D(0.0625) |
| Rozsah: | 5.9060E+02 | 8.5367E+02 | 9.7517E+02 |
| Polosuma: | 7.5415E+02 | 7.5942E+02 | 7.6423E+02 |
| Délka konců: | 0.0000E+00 | 3.6841E-01 | 5.0147E-01 |
| Šikmost: | 5.2873E-01 | 4.4424E-01 | 4.4271E-01 |
| PseudoSigma: | 4.3813E+02 | 3.7116E+02 | 3.1868E+02 |

B. Základní předpoklady

(1) TEST NORMALITY:

| | |
|--|------------|
| Tabulkový kvantil $\chi^2(1-\alpha,2)$: | 5.9915E+00 |
| χ^2 -statistika: | 2.3877E+00 |
| Závěr: Předpoklad normality přijat | |
| Vypočtená hladina významnosti: | 3.0305E-01 |

(2) TEST NEZÁVISLOSTI:

| | |
|--|------------|
| Tabulkový kvantil $t(1-\alpha/2, n+1)$: | 2.0452E+00 |
| Test autokorelace: | 1.0179E+00 |
| Závěr: Předpoklad nezávislosti přijat | |
| Vypočtená hladina významnosti: | 1.5857E-01 |

Předpoklad homogenity výběru:

| | |
|----------------------|------------|
| Aritmetický průměr: | 7.5548E+02 |
| Rozptyl: | 1.2077E+05 |
| Směrodatná odchylka: | 3.4752E+02 |

Vnitřní meze:

| | |
|-------------|-------------|
| Spodní mez: | -8.4463E+02 |
| Horní mez: | 2.3561E+03 |

(3) DETEKCE ODLEHLÝCH BODŮ:

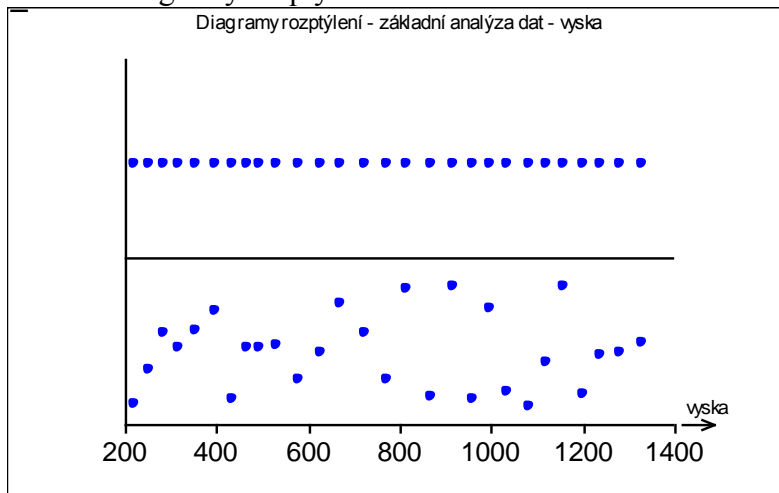
Ve výběru nejsou odlehlé body

C. Porovnání rozdělení

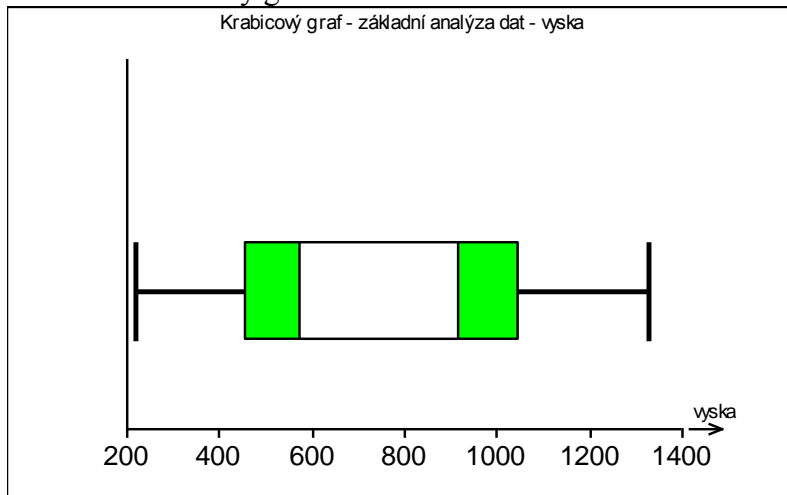
| číslo | Rozdělení | Směrnice | úsek | Korelační koeficient |
|-------|---------------|------------|------------|----------------------|
| 0 | Laplaceovo | 2.5291E+02 | 7.5548E+02 | 9.4267E-01 |
| 1 | Normální | 3.5176E+02 | 7.5548E+02 | 9.8231E-01 |
| 2 | Exponenciální | 3.4134E+02 | 4.2297E+02 | 9.0730E-01 |
| 3 | Rovnoměrné | 1.1922E+03 | 1.5939E+02 | 9.9893E-01 |
| 4 | Lognormální | 1.7913E+02 | 4.7804E+02 | 8.3125E-01 |
| 5 | Gumbelovo | 2.7190E+02 | 9.0695E+02 | 9.4936E-01 |

Normalita a nezávislost dat byly přijaty. Data nejlépe odpovídají rovnoměrnému rozdělení, neobsahují odlehlé body.

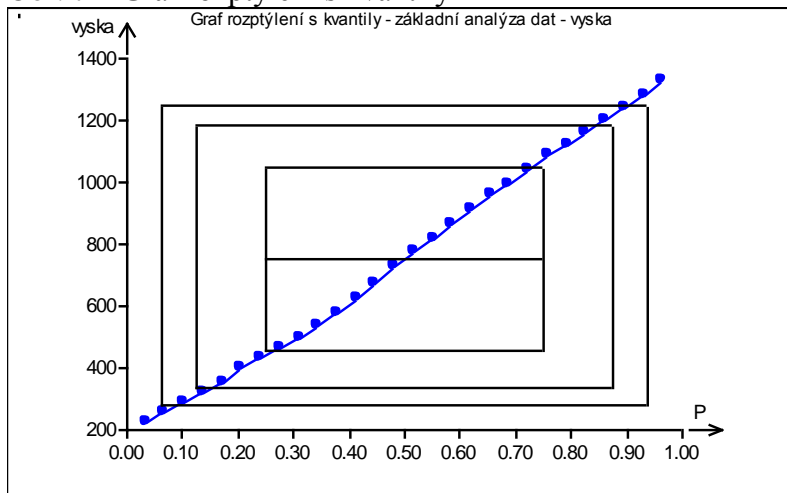
Obr. 5 – Diagramy rozptýlení



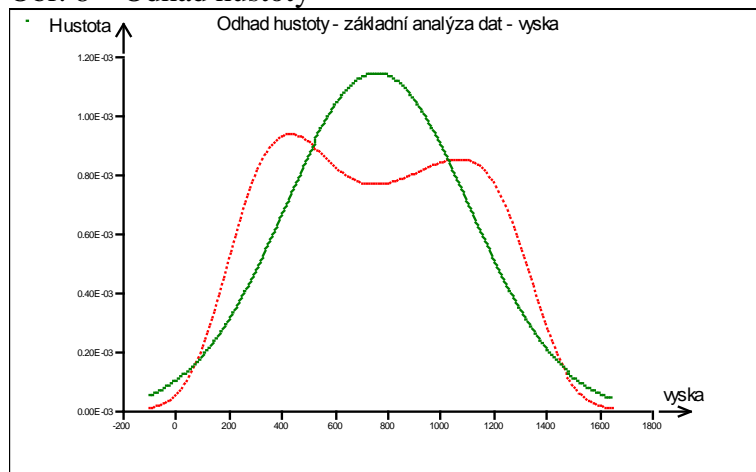
Obr. 6 – Krabicový graf



Obr. 7 – Graf rozptýlení s kvantily



Obr. 8 – Odhad hustoty



D. Odhadování parametřů

(1) BODOVÉ ODHADY PARAMETRŮ:

| Parametr | Bodový odhad | Směrodatná odchylka | Absolutní vychýlení | Relativní vychýlení[%] |
|----------|--------------|---------------------|---------------------|------------------------|
| p[1] | 1.4015E-19 | 2.8639E-17 | 2.3418E-15 | 1.6709E+06 |
| p[2] | 4.9999E+01 | 2.1125E+02 | 4.2840E+03 | 8.5680E+03 |
| p[3] | 9.6616E-01 | 1.3817E-01 | 2.3242E+00 | 2.4056E+02 |

Na dobrou podmíněnost parametrů v modelu lze usuzovat v případě, že směrodatné odchylky odhadů parametrů jsou velmi malé v porovnání s nalezenými hodnotami parametrů. V tomto případě nesplňují tuto podmínku parametry p1 a p2. Byl testován i model bez těchto dvou parametrů, statistické charakteristiky regrese se však zhoršily a model je tedy lepší se všemi třemi parametry.

(2) INTERVALOVÉ ODHADY PARAMETRŮ:

| Parametr | Bodový odhad | Poloviční délka konfidenčního int. spočtená z délky poloos | maxim |
|----------|--------------|--|----------------|
| p[1] | 1.4015E-19 | + - 8.5791E-17 | + - 8.5791E-17 |
| p[2] | 4.9999E+01 | + - 6.3283E+02 | + - 6.3283E+02 |
| p[3] | 9.6616E-01 | + - 4.1392E-01 | + - 4.1392E-01 |

(3) STATISTICKÉ CHARAKTERISTIKY REGRESE:

| | |
|--|------------|
| Reziduální součet čtverců, RSC: | 1.8912E+04 |
| Regresní rabat, D^2 [%]: | 9.9420E+01 |
| Akaikeho informační kritérium, AIC: | 1.8843E+02 |
| Průměr absolutních hodnot reziduí, MA: | 2.2869E+01 |
| Průměr relativních hodnot reziduí, MR: | 4.0155E+00 |
| Odhad reziduálního rozptylu, $s^2(e)$: | 7.5649E+02 |
| Odhad reziduální směrodatné odchylky, $s(e)$: | 2.7504E+01 |
| Odhad šikmosti reziduí, $g1(e)$: | 9.3346E-02 |
| Odhad špičatosti reziduí, $g2(e)$: | 1.7847E+00 |
| Mean error of prediction: | 9.5332E+02 |

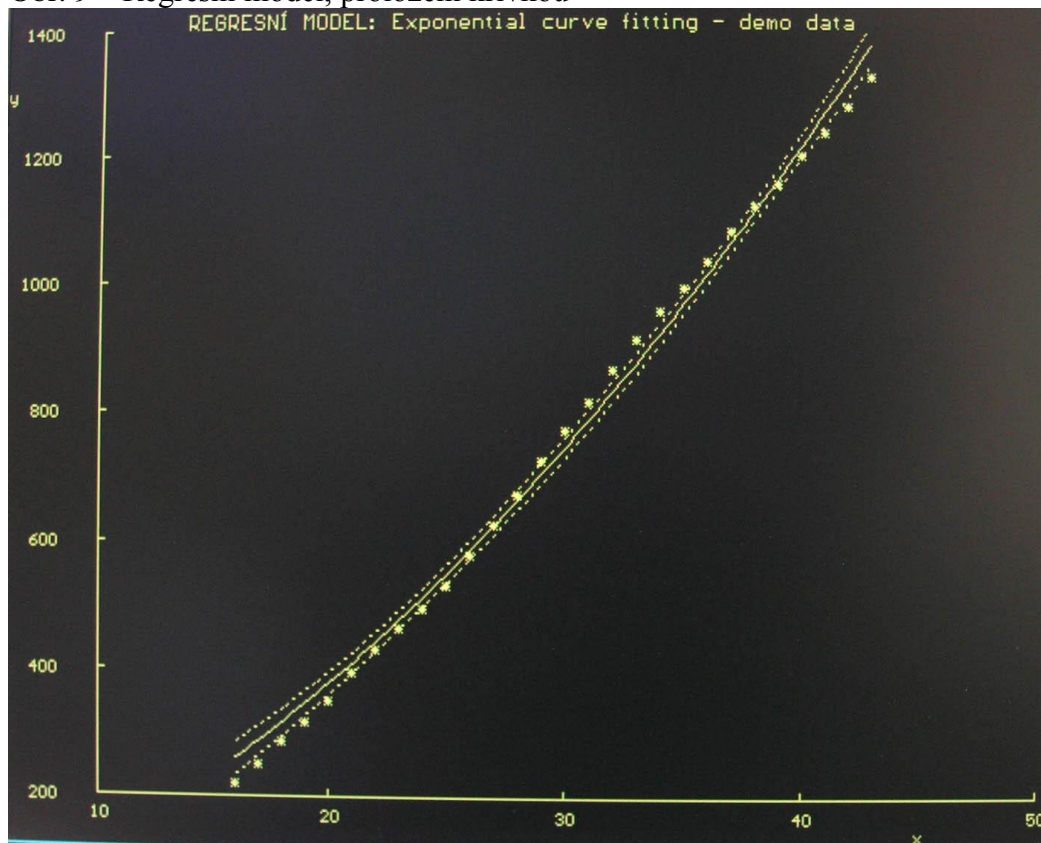
(4) INDIKACE VLIVNÝCH BODŮ:

| | Jackknife reziduum | Cookova vzdálenost | Diagonální prvky | Normalizovaná vzdálenost | Věrohodnostní vzdálenost |
|----|-----------------------|-----------------------|---------------------|-----------------------------|-----------------------------|
| i | eJ[i] | D[i] | H[i,i] | FDA | LDA |
| 1 | -1.5749E+00 | 2.0131E-01 | 2.0503E-01 | 8.4701E+03 | 3.2330E+01 |
| 2 | -1.4031E+00 | 1.3234E-01 | 1.7321E-01 | 8.4701E+03 | 3.2330E+01 |
| 3 | -1.1336E+00 | 7.1475E-02 | 1.4439E-01 | 8.4701E+03 | 3.2330E+01 |
| 4 | -9.9798E-01 | 4.5036E-02 | 1.1944E-01 | 8.4701E+03 | 3.2330E+01 |
| 5 | -8.9104E-01 | 2.9286E-02 | 9.8893E-02 | 8.4701E+03 | 3.2330E+01 |
| 6 | -4.7348E-01 | 6.9817E-03 | 8.3014E-02 | 8.4685E+03 | 3.2329E+01 |
| 7 | -2.9901E-01 | 2.3913E-03 | 7.1767E-02 | 7.7057E+03 | 3.1888E+01 |
| 8 | -3.2710E-01 | 2.5658E-03 | 6.4873E-02 | 4.1017E+03 | 2.8949E+01 |
| 9 | -5.3729E-01 | 6.5284E-03 | 6.1838E-02 | 4.8160E+02 | 1.9020E+01 |
| 10 | -5.6459E-01 | 7.2197E-03 | 6.1997E-02 | 3.1525E+09 | 9.2204E+01 |
| 11 | -2.4458E-01 | 1.4301E-03 | 6.4564E-02 | 1.8810E+08 | 7.9044E+01 |
| 12 | 1.2321E-01 | 3.8842E-04 | 6.8675E-02 | 4.8917E+03 | 2.9769E+01 |
| 13 | 3.6614E-01 | 3.6696E-03 | 7.3454E-02 | 8.4622E+03 | 3.2324E+01 |
| 14 | 8.5735E-01 | 2.0968E-02 | 7.8062E-02 | 8.4701E+03 | 3.2328E+01 |
| 15 | 1.1306E+00 | 3.7527E-02 | 8.1768E-02 | 8.4701E+03 | 3.2328E+01 |
| 16 | 1.1827E+00 | 4.2087E-02 | 8.4003E-02 | 8.4701E+03 | 3.2328E+01 |
| 17 | 1.3988E+00 | 5.7935E-02 | 8.4438E-02 | 8.4701E+03 | 3.2328E+01 |
| 18 | 1.5775E+00 | 7.0902E-02 | 8.3041E-02 | 8.4701E+03 | 3.2328E+01 |
| 19 | 1.4951E+00 | 6.1873E-02 | 8.0156E-02 | 8.4701E+03 | 3.2328E+01 |
| 20 | 1.0637E+00 | 3.1112E-02 | 7.6573E-02 | 8.4701E+03 | 3.2328E+01 |
| 21 | 8.4866E-01 | 1.9289E-02 | 7.3600E-02 | 8.4701E+03 | 3.2329E+01 |
| 22 | 7.5866E-01 | 1.5400E-02 | 7.3136E-02 | 8.4699E+03 | 3.2329E+01 |
| 23 | 3.4401E-01 | 3.4473E-03 | 7.7753E-02 | 5.9976E+03 | 3.0720E+01 |
| 24 | -1.3580E-01 | 6.3872E-04 | 9.0767E-02 | 7.9158E-01 | 4.2860E-01 |
| 25 | -5.1007E-01 | 1.1764E-02 | 1.1632E-01 | 8.2915E+03 | 3.2232E+01 |
| 26 | -1.1835E+00 | 8.7168E-02 | 1.5945E-01 | 8.4701E+03 | 3.2333E+01 |
| 27 | -1.6726E+00 | 2.5429E-01 | 2.2619E-01 | 8.4701E+03 | 3.2334E+01 |
| 28 | -2.2511E+00 | 6.9509E-01 | 3.2361E-01 | 8.4701E+03 | 3.2336E+01 |

(5) MAPA CITLIVOSTNÍ FUNKCE:

| Parametr | Relativní změna CjR(-5%) | Souhrnná citlivost Cj | Relativní změna CjR(+5%) |
|----------|--------------------------------|-----------------------------|--------------------------------|
| j | [%] | | [%] |
| 1 | 2.8706E+03 | 3.4951E+43 | 4.9346E+04 |
| 2 | 3.1113E+03 | 6.9456E+05 | 5.5413E+04 |
| 3 | 9.5695E+03 | 2.4557E+09 | 3.7024E+03 |

Obr. 9 – Regresní model, proložení křivkou

Závěr:

Byly nalezeny parametry Korfovy růstové funkce pro smrkový porost z Orlických hor. Parametr p_1 odpovídá parametru A , p_2 parametru b , p_3 parametru n .

Korfova růstová funkce má tvar šikmého protáhlého písmene S. Tohoto tvaru dosahuje u starých porostů. Experimentální data pocházejí z porostu mladého (cca 40 let), nacházejí se tedy v dolní části křivky. Tomu odpovídá i nalezené proložení experimentálních hodnot (obr. 9).