

ŠTATISTICKÁ ANALÝZA PORTFÓLIA AKCIÍ V JAZYKU R

Abstrakt

Príspevok sa zameriava na prezentáciu výpočtových možností open-source systému R pri analýze portfólia. Aplikujeme výpočtové možnosti jazyka R pre štatistickú analýzu portfólia cenných papierov, pričom využijeme najmä pokročilé grafické nástroje a taktiež postupy pre vytvorenie generátora pseudonáhodných čísel pre simulácie v kontexte so systémom náhodných premenných.

Kľúčové slová

jazyk R, Q-Q graf, histogram, časové rady, Jarqueov-Berov test, korelácia, predpoklad normality, dvojrozmerné normálne rozdelenie

Úvod

Vykonáme analýzu portfólia zloženého z akcií dvoch výrobných spoločností. Predpokladáme, že v prípade záujmu o kúpu tohto portfólia, zloženého z kombinácie týchto dvoch akcií, potrebujeme dodatočné informácie o jeho budúcom správaní, ktoré získame na základe nižšie popísanej štatistickej analýzy s využitím funkcionality jazyka R. R disponuje množstvom funkcií pre analýzu časových radov, ale v tejto analýze budeme využívať len štandardné funkcie jazyka R bez inštalácie doplnkových balíčkov (*packages*).

1 ANALYZOVANÉ ÚDAJE A ICH SPRACOVANIE

Analyzujeme údaje, ktoré pozostávajú z historických týždenných cien akcií za posledné tri kalendárne roky, t. j. 157 týždňov (hodnôt). Ceny akcií spoločností budeme označovať dvoma premennými Tyres resp. Liquor, podľa charakteru vyrábanej komodity. Údaje načítame pomocou príkazu (musia byť uložené vo formáte csv)

```
Tyresdata<-read.csv2("D:/Tyresdata.csv",header=FALSE)
Liquordata<-read.csv2("D:/Liquordata.csv",header=FALSE)
```

```
[1] 307 315 316 314 324 310 311 295 278 295 318 343 342 323 328 303 309 307
[19] 315 296 313 316 317 306 307 326 330 330 333 341 337 353 356 359 349 351
[37] 359 360 363 342 337 334 352 357 360 368 363 366 366 365 381 401 401 421
[55] 422 425 417 427 436 440 432 406 401 420 420 424 416 403 400 392 391 390
[73] 406 415 429 420 415 420 417 445 447 449 447 450 460 470 495 507 518 516
[91] 522 524 484 497 490 500 464 458 446 450 471 485 486 501 506 502 494 497
[109] 465 478 490 496 517 506 497 483 474 495 499 483 477 481 474 479 431 438
[127] 431 436 434 453 442 445 461 463 481 490 470 480 497 507 503 508 485 492
[145] 490 519 506 539 542 553 558 562 532 510 512 504 474
```

```

[1] 781 784 757 741 728 726 743 746 768 752 758 754 779 777 780
[16] 815 791 779 802 797 800 860 873 854 855 846 824 833 838 851
[31] 827 847 859 853 926 935 952 962 958 943 938 949 1000 1003 1018
[46] 1022 1026 1019 1037 1026 999 1011 994 1036 1030 1028 1005 1006 1005 970
[61] 983 984 980 996 975 976 987 1008 1057 1054 1040 1045 1057 1101 1096
[76] 1094 1108 1102 1104 1105 1092 1098 1113 1076 1060 1054 1054 1057 1078 1077
[91] 1091 1093 1079 1086 1064 1134 1167 1217 1155 1171 1174 1213 1218 1250 1329
[106] 1350 1334 1318 1315 1310 1368 1370 1389 1420 1377 1366 1424 1455 1475 1478
[121] 1458 1430 1409 1407 1414 1409 1402 1386 1392 1391 1445 1448 1462 1474 1482
[136] 1495 1525 1547 1538 1436 1465 1454 1460 1476 1521 1588 1581 1587 1539 1574
[151] 1537 1518 1530 1500 1554 1532 1498

```

Nakoľko jazyk R pracuje s údajmi, ktoré sú usporiadané vektorovo alebo vo formáte matice, tieto transformuje príkazom

```

Tyres<-as.matrix(Tyresdata)
Liquor<-as.matrix(Liquordata)

```

Pri písaní príspevku sme použili údaje podľa [2]. Používateľ však môže zvoliť akékoľvek relevantné reálne údaje resp. hodnoty je možné aj nasimulovať pomocou funkcie `sample`.

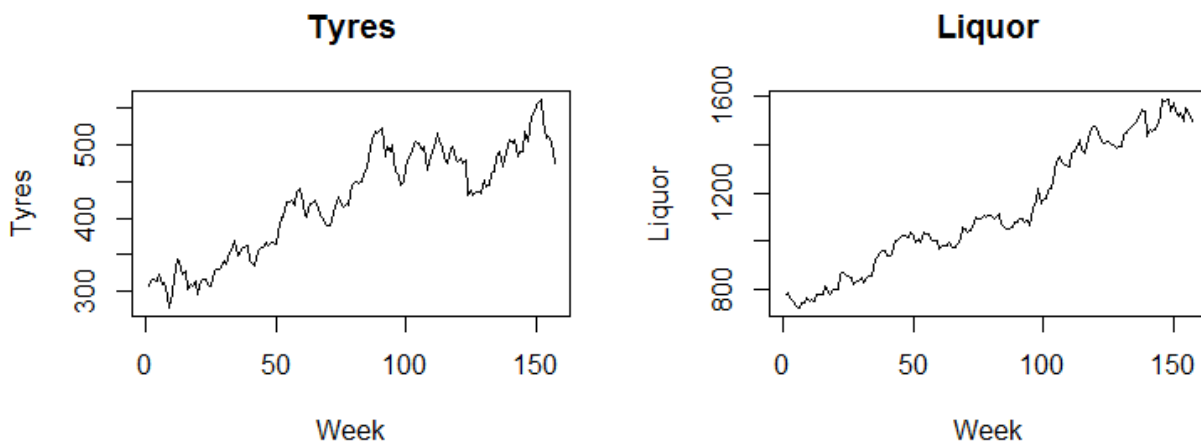
2 GRAFICKÁ ANALÝZA PORTFÓLIA

Pre zobrazenie priebehu vývoja cien akcií využijeme graf typu “l”, potom na základe nižšie uvedených príkazov dostávame výstup, ktorý zobrazuje obrázok 2.1.

```

par(mfrow=c(1,2))
plot(Tyres, main="Tyres", xlab="Week", type="l")
plot(Liquor, main="Liquor", xlab="Week", type="l")

```



Obr. 2.1

Označovanie osí v rámci výstupov (obrázkov) budeme uvádzať v anglickom jazyku. Funkcia `par` zabezpečuje voliteľné vizuálne zobrazenie jednotlivých grafov, v tomto prípade 2:1.

Označme S_t , $t = 1, 2, \dots$, cenu akcie, potom aritmetický výnos môžeme vyjadriť ako

$$r_t = \frac{S_{t+1} - S_t}{S(t)}$$

My sa však budeme zaoberať logaritmi výnosov, teda geometrickým výnosom (predpoklad *Blackovho-Scholesovho modelu*, ktorý je založený na geometrickom *Brownovom pohybe*). Využijeme teda hlavné vlastnosti, že logaritmy výnosov

$$r_{t_L} = \log \left(\frac{S_t}{S_{t-1}} \right)$$

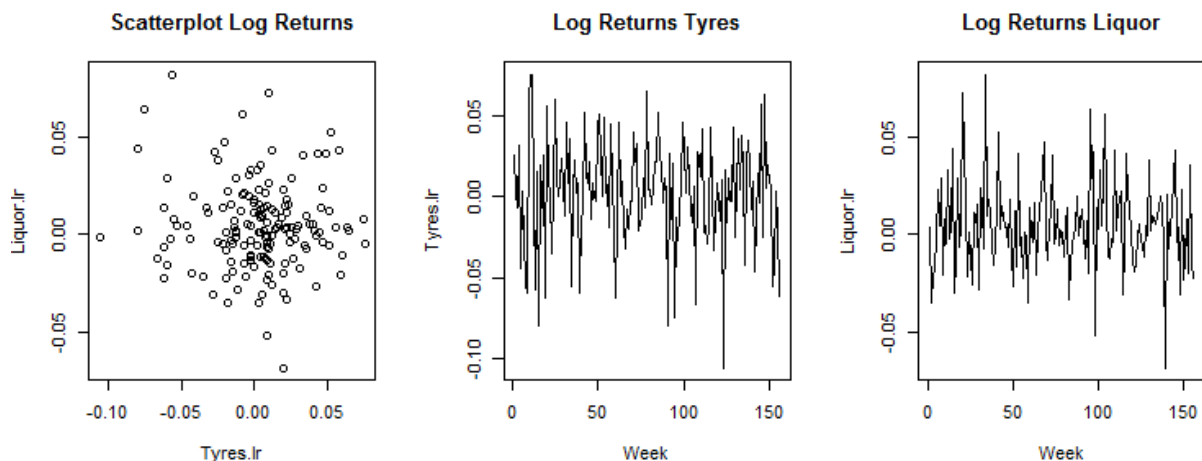
sú nezávislé a normálne rozdelené.

Na základe skutočnosti, že `Tyres [-1]` je vektor `Tyres` bez jeho prvého člena, môžeme vypočítať logaritmy výnosov bez použitia cyklu *for*. Potom tieto premenné označíme ako `Tyres.lr` resp. `Liquor.lr`

```
Tyres.lr<-log(Tyres[-1]/Tyres[-length(Tyres)])
Liquor.lr<-log(Liquor[-1]/Liquor[-length(Liquor)])
```

Pre vygenerovanie grafov logaritmov výnosov (obr. č. 2.2) využijeme príkazy

```
par(mfrow=c(1,3))
plot(Tyres.lr, Liquor.lr, main="Scatterplot Log Returns")
plot(Tyres.lr, main="Log Returns Tyres", xlab="Week", type="l")
plot(Liquor.lr, main="Log Returns Liquor", xlab="Week", type="l")
```

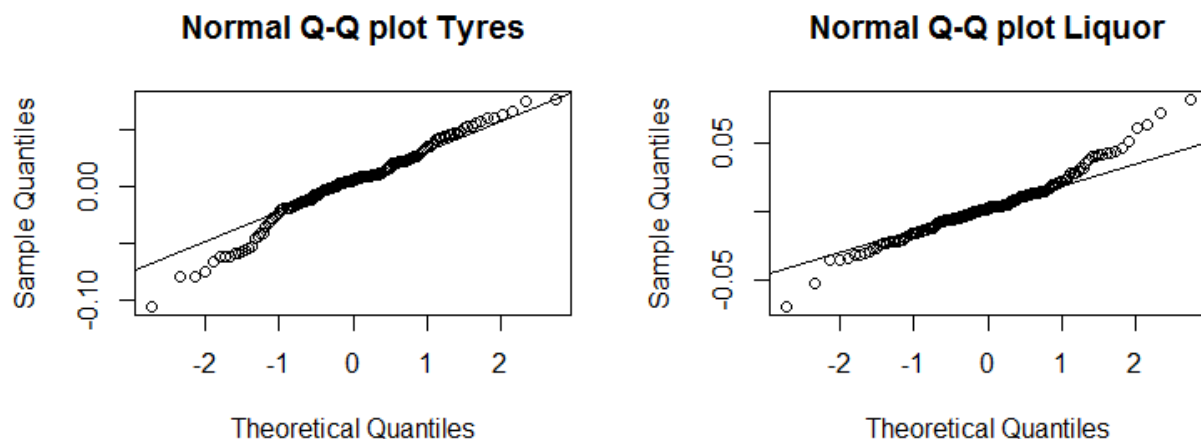


Obr. 2.2

Q-Q graf (Q-Q plot) je vizuálnym (často však subjektívnym) grafickým testom pre posúdenie normality. Porovnáva kvantily analyzovaného súboru hodnôt s teoretickými kvantilmi (normálneho rozdelenia).

```
par(mfrow=c(1,2))
```

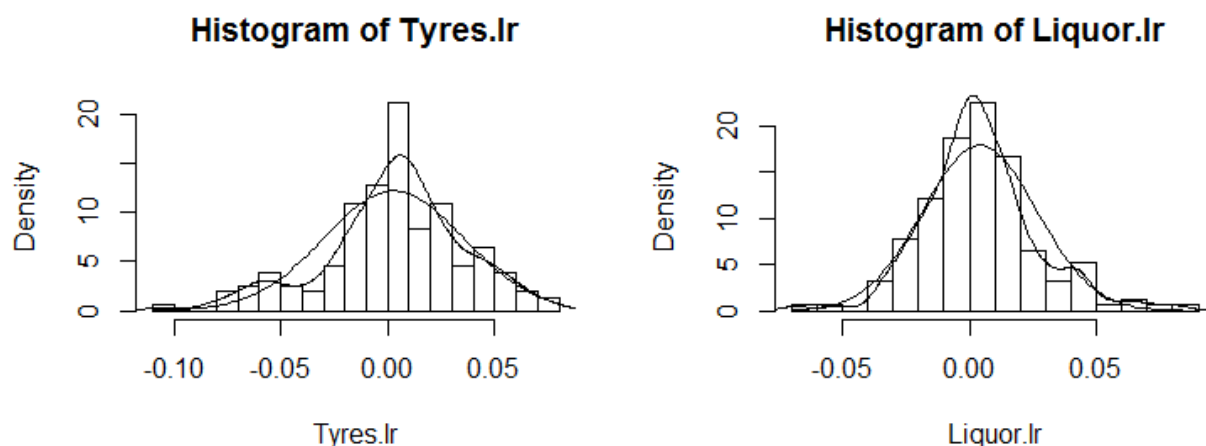
```
qqnorm(Tyres.lr,main="Normal Q-Q plot Tyres")
qqline(Tyres.lr)
qqnorm(Liquor.lr,main="Normal Q-Q plot Liquor")
qqline(Liquor.lr)
```



Obr. 2.3

V našom prípade (podľa obr. č. 2.3) možno zamietnuť hypotézu, že logaritmy výnosov sa riadia normálnym rozdelením v oboch prípadoch (hodnoty sa na začiatku a na konci nie najtesnejšie kumulujú okolo diagonály)

Ďalším často uvádzaným grafickým testom normality je *histogram*. Obrázok 2.4 zobrazuje histogram preložený krivkou hustoty normálneho rozdelenia (so strednou hodnotou a štandardnou odchýlkou jednotlivých logaritmov výnosov akcií). Parameter `breaks` funkcie `hist` určuje nastavenie počtu tried.



Obr. 2.4

```
par(mfrow=c(1,2))
hist(Tyres.lr, prob=T, breaks=21)
curve(dnorm(x, mean=mean(Tyres.lr), sd=sd(Tyres.lr)), add=T)
lines(density(Tyres.lr))
hist(Liquor.lr, prob=T, breaks=21)
curve(dnorm(x, mean=mean(Liquor.lr), sd=sd(Liquor.lr)), add=T)
lines(density(Liquor.lr))
```

Funkcia `density` predstavuje jadrový odhad hustoty pravdepodobnosti (*kernel density estimate*). Tieto odhady patria medzi neparametrickej odhady, tzn. netreba vopred predpokladať aký tvar má odhadovaná hustota. Vzhľadom na náročný teoretický aparát sa ďalej touto problematikou nebudem zaoberať (používateľ môže nájsť ďalšie informácie po spustení príkazu `??density`).

Rovnako môžeme pomocou grafickej analýzy napríklad skúmať *volatilitu*, teda dôležitú mieru kolísania hodnoty aktíva alebo jeho výnosovej miery, ktorá rovnako vyjadruje mieru rizika investície do určitého aktíva. Volatilita je tiež termínom pre vyjadrenie nestálosti či zmeny, pričom u finančných aktív rastie volatilita s odmocninou časového úseku, na ktorom je meraná. Na tieto analýzy slúžia ďalšie funkcie jazyka R. Pre viac informácií pozri problematiku týkajúcu sa volatility modelov cien akcií.

3 ANALYTICKÁ ANALÝZA PORTFÓLIA

V tejto časti sa najskôr budeme zaoberať ďalšími testami normality. Zo širokej ponuky testov normality zvolíme *Jarqueov-Berov test*. Priamo uvádzame samostatný program v jazyku R, ktorý možno využiť na aplikáciu tohto testu pre analyzované údaje. Teoretické predpoklady testu možno nájsť v príslušnej odbornej literatúre. Testovaciu charakteristiku v ktorej vystupuje šikmosť a špicatosť analyzovaného súboru (viď program) označíme ako JB. Vstupnými hodnotami sú opätovne `Tyres.lr` resp. `Liquor.lr`.

```
x <- Tyres.lr-mean(Tyres.lr)
m2 <- mean (x^2); m3 <- mean(x^3); m4 <- mean (x^4)
S2 <- m3^2/m2^3; K<- m4/m2^2 - 3
JB <- length(x)/6*(S2 + K^2/4)
p.value <- 1-pchisq (JB, df=2)
```

Pre `Tyres` dostávame výstup (9.30528, 0.00954). Rovnako napr. pre `Liquor` dostávame (18.88978, 0.000079). S pravdepodobnosťou 95 % zamietame normalitu logaritmov výnosov `Tyres` aj `Liquor` (prijímame alternatívnu hypotézu).

V ďalších úvahách nebudeme predpokladať len marginálnu normalitu, ale v skutočnosti budeme pre logaritmy výnosov uvažovať, že sa riadia dvojrozmerným normálnym rozdelením.

Ak náhodné premenné X, Y (označenie pre náš prípad podľa názvov jednotlivých akcií $T \equiv X, L \equiv Y$) majú marginálne normálne rozdelenie, potom pre ich charakteristiky platí

$$E(X) = \mu_X, \sigma(X) = \sigma_X, E(Y) = \mu_Y, \sigma(Y) = \sigma_Y$$

a tieto jednoducho vypočítame pomocou funkcií `mean`, `sd` a tiež `corr`

```
Tyres.lr.mean <- mean(Tyres.lr)
Tyres.lr.sd <- sd(Tyres.lr)
Liquor.lr.mean <- mean(Liquor.lr)
Liquor.lr.sd <- sd(Liquor.lr)
Tyres.Liquor.lr.corr <- cor(Tyres.lr, Liquor.lr)
```

potom dostávame postupne výstup piatich hodnôt

0.00278, 0.03226, 0.00418, 0.02235, 0.06142

kde posledná hodnota predstavuje koeficient korelácie $r_{X,Y}$ systému (X, Y).

Predpokladajme, že chceme kúpiť portfólio rovnakej časti Tyres akcií za bežnú cenu napr. 474 peň. j. a Liquor akcií za cenu 1498 peň. j. a zaujíma nás napríklad ich budúca výkonnosť v horizonte dvoch kalendárnych rokov (t. j. 104 týždňov). Potom týždenné logaritmické výnosy akcií systému (X, Y) pre túto periódu

$$(X_i, Y_i), i = 158, \dots, 261$$

sa riadia dvojrozmerným normálnym rozdelením, s parametrami vypočítanými vyššie, a sú nezávislé pre rôzne i . Môžeme písať, že

$$X = X_{158} + \dots + X_{261}$$

$$Y = Y_{158} + \dots + Y_{261}$$

a pre náhodnú premennú S platí

$$S = 474e^X + 1498e^Y$$

Charakteristiky, kovarianciu systému (X, Y) a koeficient korelácie $r_{X,Y}$ pre periódu 104 týždňov, vypočítame pomocou nasledovných funkcií

```
Periods <- 104
mean.X <- Periods * Tyres.lr.mean
mean.Y <- Periods * Liquor.lr.mean
sd.X <- sqrt(Periods * Tyres.lr.sd^2)
sd.Y <- sqrt(Periods * Liquor.lr.sd^2)
cov.XY <- Periods * Tyres.Liquor.lr.corr * Tyres.lr.sd *
Liquor.lr.sd
r.XY <- cov.XY / sd.X / sd.Y
```

Výstup pre kovarianciu a koeficient korelácie systému (X, Y) podľa

$$r = \frac{\text{cov}(X, Y)}{\sigma_X \cdot \sigma_Y} \Leftrightarrow \text{cov}(X, Y) = r \cdot \sigma_X \cdot \sigma_Y$$

a definovanú periódu je

0.0046, 0.06142

Na základe vyššie uvedeného, systém náhodných premenných (X, Y) má dvojrozmerné normálne rozdelenie a náhodná premenná S je súčtom závislých lognormálne rozdelených náhodných premenných. Výpočet distribučnej funkcie resp. kvantilov náhodnej premennej S je však mimoriadne náročné. Jednou z možností je simulovanie hodnôt tejto náhodnej

premennej. Pre generovanie hodnôt viacrozmerného normálneho rozdelenia možno v jazyku R využiť napr. funkciu `mvrnorm` (balíček `MASS`).

Problém môžeme však redukovať aj využitím generovania hodnôt jednorozmerných nezávislých normovane normálne rozdelených náhodných premenných, ktoré označme U , V , pre všetky i . Potom (s využitím metódy Choleského rozkladu) pre koeficient korelácie platí

$$r(U, U + \alpha V) = \frac{\text{cov}(U, U + \alpha V)}{\sigma_U \cdot \sigma_{U + \alpha V}} = \frac{1}{\sqrt{1 + \alpha^2}}$$

ak $r > 0$

$$\alpha = \sqrt{\frac{1}{r^2} - 1}, \quad \alpha \in \mathbb{R}$$

Potom náhodná premenná $W = r \cdot (U + \alpha V)$ má rovnako normované normálne rozdelenie a pre prípad, ak $r \leq 0$, platí $W = V$. Pre situáciu, ktorú sme naformulovali vyššie potom platí

$$S^* = 474e^{X^*} + 1498e^{Y^*}$$

kde

$$X^* = E(X) + U\sqrt{D(X)}$$

$$Y^* = E(Y) + W\sqrt{D(Y)}$$

a platí

$$S^* \sim S$$

Kód nasledujúceho programu (podľa [2]) je teda generátorom pseudonáhodných čísel, ktorý generuje 1000 hodnôt náhodnej premennej S so známymi parametrami systému (X, Y) , ktorý má dvojrozmerné normálne rozdelenie. Využili sme štandardnú funkciu `rnorm`.

```
set.seed(2525) ## initialize the generator
U <- rnorm (1000); V <- rnorm (1000)
alpha <- sqrt (1/r.XY^2-1) * sign(r.XY)
X <- mean.X + sd.X *U
Y <- mean.Y + sd.Y *(U + alpha * V) * r.XY
S <- Tyres[length(Tyres)] * exp(X) + Liquor[length(Liquor)] *
exp(Y)
```

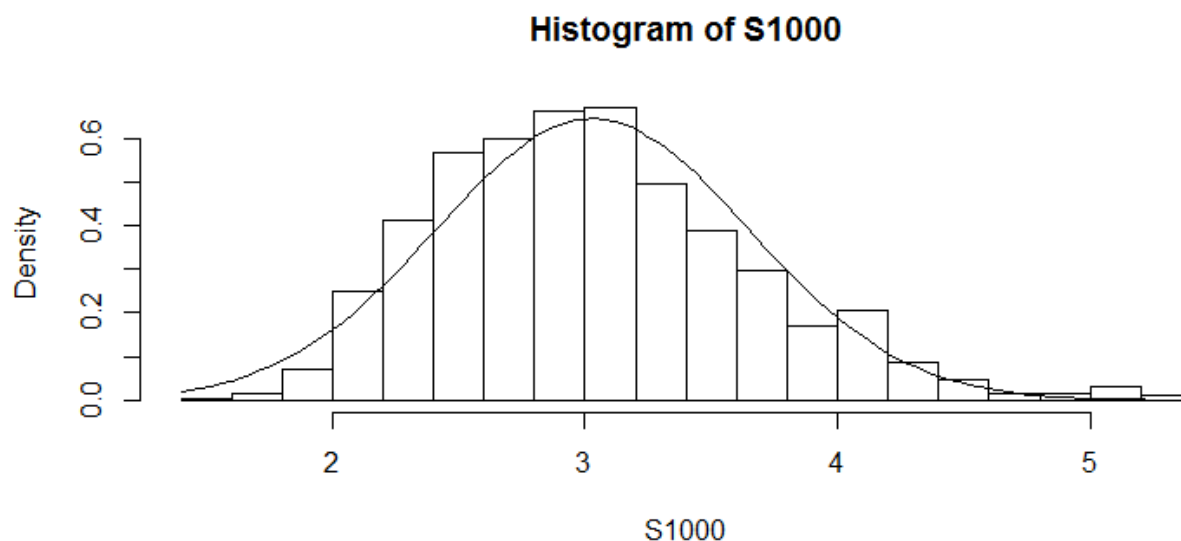
Časť výstupu vygenerovaných hodnôt náhodnej premennej S je nasledovný

```
[1] 2895.531 3559.751 3490.921 2345.381 3191.637 2751.059 2676.651 2381.337
[9] 2984.299 2447.416 3466.974 3436.462 2375.028 2362.195 3347.027 3476.856
[17] 2221.862 3352.707 3544.526 3683.836 2635.199 3241.413 2654.776 2343.645
[25] 3065.252 2857.519 2829.346 4392.884 2269.309 2997.395 4456.613 3100.836
[33] 2716.946 2568.026 3543.810 2959.699 3026.186 3049.025 2497.242 3225.004
[41] 3490.510 2293.856 2762.288 3351.175 3614.784 3312.997 3654.899 2592.343
[49] 3034.216 1967.872 3867.123 3444.299 2560.822 3144.356 3585.653 3413.243
[57] 2884.011 3128.587 2394.615 2397.573 4670.996 3793.242 4283.365 2733.163
[65] 2791.947 2858.460 2700.248 2325.359 3709.467 3266.619 2437.770 2917.315
[73] 2509.756 2026.783 2420.246 2174.615 2624.140 3447.669 3423.768 2285.464
```

```
[81] 2661.776 2628.434 2448.423 3063.918 2768.999 3567.311 2723.243 2851.457
[89] 3572.612 2997.512 3616.474 2817.176 2385.286 3756.942 2256.518 3590.286
[97] 3165.379 3329.213 2724.249 3211.277 3951.678 3487.243 2566.680 2869.346
[105] 3276.419 2141.070 2795.334 2697.397 2092.781 2965.996 2132.946 3534.743
```

Obrázok 3.1 znázorňuje histogram hodnôt náhodnej premennej S (s modifikáciou hodnôt ako $S/1000$) preložený hustotou normálneho rozdelenia. Vygenerujeme ho pomocou nasledovných funkcií

```
par(mfrow=c(1,1)); S1000 <- S/1000
hist(S1000, prob=T, breaks=21)
curve(dnorm(x,mean=mean(S1000),sd=sd(S1000)),add=T)
```



Obr. 3.1

Nakoniec určíme niektoré kvantily náhodnej premennej S ako

```
pr <-c(2.5, 5, 10, 25, 50, 75, 90, 97.5, 99, 99.5)/100
S.pr <- S/( Tyres[length(Tyres)] + Liquor[length(Liquor)])
round(100*quantile(S.pr,pr))
```

a dostávame vstup

2.5%	5%	10%	25%	50%	75%	90%	97.5%	99%	99.5%
103	111	119	134	152	172	190	223	238	245

Záver

Z dôvodu rozsiahleho pojmového aparátu a vystupujúcich premenných sme do tabuľky č. 1 usporiadali všetky definované premenné (objekty) a do tabuľky č. 2 uviedli relevantné numerické výsledky našej analýzy.

Tab. 1: Prehľad premenných (objektov)

Vstupné údaje $n_1 = n_2 = 157$	Tyres Liquor
Logaritmy výnosov akcií	Tyres.lr Liquor.lr
Grafická analýza portfólia	plot qqnorm hist curve
Jarqueov-Berov test normality	JB p.value
Charakteristiky(X, Y)	Tyres.lr.mean Tyres.lr.sd Liquor.lr.mean Liquor.lr.sd
$r_{X,Y}$	Tyres.Liquor.lr.corr
Charakteristiky(X, Y)*	mean.X mean.Y sd.X sd.Y
cov*(X, Y)	cov.XY
$r_{X,Y}^*$	r.XY
Hodnoty náhodnej premennej S (simulácia)	S
Kvantily náhodnej premennej S	round(100*quantile(S.pr,pr))

Zdroj: vlastné spracovanie, s formalizáciou podľa [2]

Skúsený analytik môže následne výsledky (najmä hodnoty kvantilov náhodnej premennej S a grafy podľa obr. č. 2.1 až 2.4) adekvátne využiť pre predikciu budúceho správania portfólia, ak ceny akcií v budúcnosti sa budú riadiť modelom, ktorý sme odhadli na základe vstupných údajov (historických cien) a predpokladov transakcie.

Tab. 2: Prehľad výsledkov riešenia

	Tyres	Liquor
JB	9.30528	18.88978
p – value (JB)	0.00954	0.00008
$\mu_X ; \mu_Y$	0.00278	0.00418
$\sigma_X ; \sigma_Y$	0.03226	0.02235
$r_{X,Y}^*$	0.06142	
cov*(X, Y)	0.00460	
$s_{0,25}$	134	
$s_{0,5} = Me(S)$	152	
$s_{0,75}$	172	

Zdroj: vlastné spracovanie

Jazyk R teda poskytuje (aj bez inštalácie doplnkových balíčkov) prostredníctvom zabudovaných funkcií prostredie, kde bolo možné spracovať všetky výpočtové požiadavky. Výhodou (a z dôvodu náročnosti niekedy aj nevýhodou) oproti komerčným softvérom je skutočnosť, že používateľ v zdrojovom kóde reálne vidí priebeh celého výpočtu. Špičkový grafický výstup a množstvo výpočtových modifikácií stavia open-source systém R do popredia štatistických nástrojov nie len v akademickej a vedeckej sfére.

Predstavili sme využitie jazyka R pri analýze portfólia ako riešiteľského nástroja v oblasti manažmentu rizík a rozhodovania na trhu s finančnými nástrojmi. Analyzovali sme historické údaje o týždenných cenách akcií dvoch spoločností za tri kalendárne roky. Budúce správanie portfólia sme sa snažili predikovať najskôr na základe grafickej analýzy, kde sme zobrazili grafický vývoj skutočných cien aj logaritmov výnosov akcií, *Q-Q graf* a histogram logaritmov výnosov. *Graf Q-Q* je subjektívnym testom pre posúdenie normality sledovaného súboru, preto sme ju rovnako overovali aj napr. Jarque-Berovým testom. Následne sme skúmali situáciu, pri ktorej sme predpokladali, že náhodné premenné (logaritmy výnosov) môžeme považovať za systém náhodných premenných, ktorý sa riadi dvojrozmerným normálnym rozdelením. Na základe vytvoreného generátora pseudonáhodných čísel sme pomocou simulácií hodnôt náhodnej premennej, ktorú sme definovali na základe viacerých predpokladov, získali podstatné vlastnosti tohto modelu.

Použitá literatúra

- [1] HUŤKA, V. 2005. *Teória pravdepodobnosti 2*. Bratislava : Vydavateľstvo EKONÓM, 2005. ISBN 80-225-2040-3.
- [2] KAAS, R., GOOVAERTS, M., DHAENE, J., DENUIT, M. 2008. *Modern actuarial risk theory using R*. Berlin : Springer, 2008. ISBN 978-3-642-03407-7
- [3] PÁLEŠ, M. 2014. *Využitie software pri výučbe predmetov z oblasti aktuárstva*. MITAV 2014. Brno : Univerzita obrany v Brně. ISBN 978-80-7231-961-9
- [4] ŠOLTÉS, E. 2008. *Regresná a korelačná analýza s aplikáciami*. Bratislava : Iura Edition, 2008. ISBN 978-80-8078-163-7.

Kontaktné údaje

Páleš, Michal, Ing., PhD., Katedra matematiky a aktuárstva, Fakulta hospodárskej informatiky, Ekonomická univerzita v Bratislave, Dolnozemska cesta 1, 852 35 Bratislava, tel. +421 2/672 95 841, e-mail: pales.euba@gmail.com