VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA STROJNÍHO INŽENÝRSTVÍ
FACULTY OF MECHANICAL ENGINEERING

ÚSTAV MATEMATIKY
INSTITUTE OF MATHEMATICS

KORELAČNÍ KOEFICIENTY A OBLASTI JEJICH POUŽITÍ

CORRELATION COEFFICIENTS AND THEIR FIELD OF APPLICATION

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE
AUTHOR
Matej Rudžík

VEDOUCÍ PRÁCE
SUPERVISOR
doc. RNDr. Libor Žák, Ph.D.

BRNO 2018

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA STROJNÍHO INŽENÝRSTVÍ

Zadání bakalářské práce

Ústav:	Ústav matematiky
Student:	**Matej Rudžík**
Studijní program:	Aplikované vědy v inženýrství
Studijní obor:	Matematické inženýrství
Vedoucí práce:	**doc. RNDr. Libor Žák, Ph.D.**
Akademický rok:	2017/18

Ředitel ústavu Vám v souladu se zákonem č.111/1998 o vysokých školách a se Studijním a zkušebním řádem VUT v Brně určuje následující téma bakalářské práce:

Korelační koeficienty a oblasti jejich použití

Stručná charakteristika problematiky úkolu:

Cílem práce je porovnat parametrické i neparametrické korelační koeficienty pro různé rozložení dat.

Cíle bakalářské práce:

1. Nastudování různých typů parametrických korelačních koeficientů.
2. Nastudování různých typů neparametrických korelačních koeficientů.
3. Zvolení vhodných dat pro použití regresních koeficientů.
4. Nad daty provést korelační analýzu.
5. Vyhodnocení výsledků.

Seznam doporučené literatury:

ANDĚL, Jiří. Základy matematické statistiky. 3. opr. vyd. Praha: Matfyzpress, 2011. ISBN 978-8--7378-001-2.

Termín odevzdání bakalářské práce je stanoven časovým plánem akademického roku 2017/18

V Brně, dne

L. S.

prof. RNDr. Josef Šlapal, CSc.
ředitel ústavu

doc. Ing. Jaroslav Katolický, Ph.D.
děkan fakulty

Abstrakt

Táto bakalárska práca sa zaoberá rôznymi typmi korelačných koeficientov a ich vlastnosťami v závislosti od rozloženia dát. Teoretická časť práce obsahuje definície jednotlivých typov korelačných koeficientov a testov súvisiacich s touto problematikou. Praktická časť porovnáva chovanie parametrických a neparametrických korelačných koeficientov v rôznych prípadoch. Toto porovnanie je realizované v programe MATLAB.

Abstract

This bachelor's thesis deals with different types of correlation coefficients and their properties depending on the data layout. The theoretical part of this work contains definitions of various types of correlation coefficients and tests related with this issue. The practical part compares the behaviour of parametric and nonparametric correlation coefficients in different situations. This comparison is realised in program MATLAB.

Kľúčové slová

korelácia, korelačný koeficient, Pearsonov korelačný koeficient, Spearmanov korelačný koeficient

Keywords

correlation, correlation coefficient, Pearson correlation coefficient, Spearman's correlation coefficient

RUDŽÍK, M. *Korelační koeficienty a oblasti jejich použití.* Brno: Vysoké učení technické v Brně, Fakulta strojního inženýrství, 2018. 50 s. Vedoucí bakalářské práce doc. RNDr. Libor Žák, Ph.D.

Prehlasujem, že som svoju bakalársku prácu na tému *Korelační koeficienty a oblasti jejich použití* vypracoval samostatne pod vedením vedúceho bakalárskej práce, doc. RNDr. Libora Žáka, Ph.D., s použitím materiálov uvedených v zozname literatúry.

V Brne dňa 16.5.2018

Matej Rudžík

Na tomto mieste by som sa chcel poďakovať doc. RNDr. Liborovi Žákovi, Ph.D., vedúcemu mojej bakalárskej práce, za jeho odborné rady a čas strávený pri konzultácii tejto práce. Taktiež by som sa chcel poďakovať mojim rodičom za trpezlivosť a pomoc počas celého štúdia.

Obsah

Úvod

V každodennom živote sa stretávame s tým, že jednotlivé javy sú na sebe navzájom závislé a vzájomne sa ovplyvňujú. Preto často hľadáme cestu, ako účinne vyjadriť vzájomnú závislosť dvoch, prípadne viacerých javov. V oblasti matematickej štatistiky je nástrojom na skúmanie tejto závislosti korelačná analýza.

V tejto práci sa obmedzíme na závislosť dvoch veličín, ktorej sa v dvojrozmernej štatistike venuje jednoduchá korelačná analýza. Prostriedkom na vyjadrenie tejto závislosti je v korelačnej analýze korelačný koeficient. Samotný korelačný koeficient je však ovplyvnený tým, ako sú v skúmanej vzorke rozložené dáta. Práve v závislosti od rozloženia dát vo vzorke sa odvíja to, ktorý typ korelačného koeficientu je vhodné použiť. Vhodnou voľbou typu koeficientu môžeme zamedziť tomu, aby boli získané výsledky skreslené a dávali nepresný popis skúmaného vzťahu medzi veličinami.

Táto práca pozostáva z troch kapitol, v ktorých sú rozobrané základné pojmy z jednoduchej korelačnej analýzy. V prvej kapitole si zadefinujeme korelačný koeficient a uvedieme jeho základné vlastnosti. Následne uvedieme niekoľko najpoužívanejších typov korelačných koeficientov. Druhá kapitola pozostáva z teórie týkajúcej sa rôznych testov spojených s korelačnými koeficientami. Tretiu kapitolu tvorí praktická časť, ktorá sa týka spomínaného faktu, že korelácia medzi dvoma veličinami môže byť ovplyvnená rozložením dát vo vzorke. V prostredí MATLAB porovnáme chovanie rôznych korelačných koeficientov v závislosti na rôznych faktoroch, ako je rozsah súboru či rozdelenie oboch veličín vystupujúcich vo výpočte korelácie.

1 Korelačný koeficient

Ak by sme mali vo všeobecnosti vysvetliť význam slova korelácia, mohli by sme ho označiť ako hodnotu stupňa vzájomnej asociácie dvoch veličín. Skutočnosť, kedy dve veličiny spolu asociujú (korelujú) zachytáva fakt, že hodnoty jednej veličiny majú tendenciu sa vyskytovať spoločne s hodnotami druhej veličiny. Miera tejto tendencie je však prirodzene rôzna. V praxi sa môže pohybovať od neexistujúcej korelácie až po absolútnu koreláciu. Samotné konštatovanie, že dve veličiny spolu korelujú, prípadne nekorelujú, nám prirodzene nestačí. Preto bol pre meranie korelácie navrhnutý korelačný koeficient. Ten je ale ovplyvnený typom dát vo vzorke. Kvôli tomuto faktu bolo navrhnutých hneď niekoľko typov korelačných koeficientov. [5, 6]

Ešte pred samotným rozhodnutím o použití konkrétneho korelačného koeficientu potrebujeme logicky analyzovať skúmaný problém. Už len logická úvaha nám často napovie, kedy je koreláciu zbytočné merať z dôvodu, že je skreslená, prípadne vôbec neexistuje. Medzi takéto prípady patrí napríklad formálna korelácia, ktorá vzniká pri zisťovaní korelácie percentuálnych charakteristík dopĺňajúcich sa do 100 %. Ďalším faktorom, ktorý skresľuje hodnotu korelačného koeficientu, je nehomogenita študovaného súboru. Pri dvojrozmerných metódach skúmania sa môžeme často stretnúť s falošnou koreláciou medzi veličinami X a Y, ktorá vzniká v dôsledku nezohľadnenia vplyvu jednej, prípadne viacerých ďalších veličín. Je nutné poznamenať, že na základe korelácie veličín nemôžeme usudzovať ich príčinný vzťah. Teda na základe faktu, že veličiny spolu korelujú, nemôžeme vyvodiť záver o tom, že by sa navzájom podmieňovali. [5, 6]

Táto kapitola nám poslúži na zavedenie základných pojmov nutných pre definíciu samotného korelačného koeficientu. Následne sa oboznámime s rôznymi druhmi korelačných koeficientov.

1.1 Definícia korelačného koeficientu

Ešte pred samotnou definíciou korelačného koeficientu si zavedieme pojmy potrebné pre jeho definovanie.

Kovariancia: Rovnako ako korelácia, aj kovariancia patrí ku charakteristikám, ktoré číselne vyjadrujú lineárnu závislosť medzi veličinami X a Y. Definujeme ju nasledujúcim vzťahom:

$$C(X,Y) = E([X - E(X)][Y - E(Y)]). \tag{1.1}$$

Pre praktický výpočet sa však častejšie používa vzťah:

$$C(X,Y) = E(X \cdot Y) - E(X) \cdot E(Y), \tag{1.2}$$

kde $E(X)$ je stredná hodnota náhodnej veličiny X definovaná vzťahom:

$$E(X) = \int_{\omega \epsilon \Omega} X(\omega) dP(\omega). \tag{1.3}$$

Poznámka: Vo vzťahu pre výpočet strednej hodnoty náhodnej veličiny X je $P(\omega)$ pravdepodobnostná miera určujúca rozdelenie náhodnej veličiny X.

Kovariancia nadobúda hodnoty z intervalu $(-\infty, \infty)$. Svojím znamienkom nám dáva informáciu o tom, či je vzťah medzi nami skúmanými veličinami priamy alebo nepriamy.

Pokiaľ sa nadpriemerné (podpriemerné) hodnoty veličiny X združujú s nadpriemernými (podpriemernými) hodnotami veličiny Y, bude kovariancia kladná. Teda spoločne s nárastom hodnôt jednej veličiny rastú aj hodnoty druhej veličiny, a naopak s poklesom hodnôt jednej veličiny klesajú aj hodnoty druhej veličiny. Pokiaľ sa nám združujú nadpriemerné (podpriemerné) hodnoty jednej veličiny s podpriemernými (nadpriemernými) hodnotami druhej veličiny, bude kovariancia záporná. To znamená, že s nárastom hodnôt jednej veličiny klesajú hodnoty druhej veličiny a naopak. Nulová kovariancia nám dáva informáciu o tom, že medzi veličinami neexistuje lineárna závislosť. Nemusí to však ešte znamenať, že medzi znakmi neexistuje iný nelineárny vzťah. [2]

Korelácia: Keďže nám kovariancia nehovorí nič o sile lineárneho vzťahu medzi veličinami, je prirodzené hľadať efektívnejší prostriedok na vyjadrenie tohto vzťahu. Tým je korelačný koeficient, ktorý definujeme vzťahom:

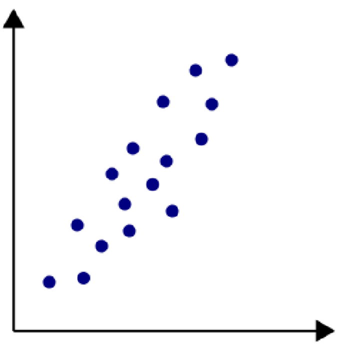
$$\rho(X,Y) = E\left(\frac{X - E((X)}{\sqrt{D(X)}} \cdot \frac{Y - E(Y)}{\sqrt{D(Y)}}\right) = \frac{C(X,Y)}{\sqrt{D(X)} \cdot \sqrt{D(Y)}}, \tag{1.4}$$

kde $D(X)$ je rozptyl náhodnej veličiny X definovaný vzťahom:

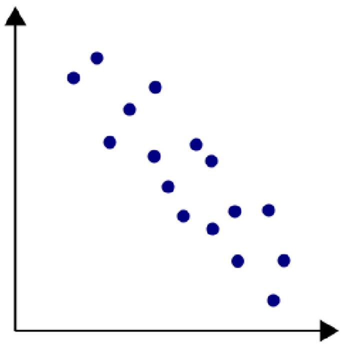
$$D(X) = E([X - E(X)]^2). \tag{1.5}$$

Poznámka: Výraz $\sqrt{D(X)}$ vo vzorci (1.4) pre výpočet korelačného koeficientu sa nazýva aj smerodajná odchýlka $\sigma(X)$. Platí: $\sigma(X) = \sqrt{D(X)}$. Výpočet korelačného koeficientu má teda zmysel v prípade, že sú obe smerodajné odchýlky nenulové.

Korelačný koeficient je bezrozmerné číslo, nezávislé na tom, v akých jednotkách boli merané náhodné veličiny X a Y. Nadobúda hodnoty z intervalu $\langle -1,1 \rangle$. Kladné hodnoty koeficientu znamenajú kladnú koreláciu, záporné hodnoty znamenajú zápornú koreláciu. Čím viac sa hodnota korelačného koeficientu blíži k hodnote 1 (respektíve -1), tým silnejší je lineárny vzťah medzi skúmanými veličinami X a Y. Hodnota korelačného koeficientu blížiaca sa k nule nám indikuje slabnutie tohto lineárneho vzťahu. Nemôžeme však usúdiť, že v prípade jeho nulovej hodnoty sú veličiny X a Y nezávislé. Môže medzi nimi existovať iná ako lineárna závislosť. [2, 3]



Obrázok 1.1: Kladná korelácia



Obrázok 1.2: Záporná korelácia

Uvedieme niekoľko základných vlastností korelačného koeficientu:

1. $|\rho(X,Y)| \leq 1$
2. $\rho(X,Y) = \rho(Y,X)$
3. $|\rho(X,Y)| = 1,$ pokiaľ $P(Y = aX + b) = 1,\ a,b \in \mathbb{R}, a \neq 0.$

Problémom pri vzťahu (1.4) pre výpočet korelačného koeficientu je podmienka znalosti simultánneho rozdelenia vektoru $(\mathbb{X}, \mathbb{Y})$ pozostávajúceho z nami skúmaných náhodných veličín X a Y. V praxi však k tejto situácii dôjde málokedy. Musíme sa preto obmedziť na náhodný výber $\left(\binom{X_1}{Y_1}, \dots, \binom{X_n}{Y_n}\right)$ o rozsahu n z dvojrozmerného rozdelenia, teda na výber z navzájom nezávislých náhodných veličín s rovnakým rozložením $L(\vartheta)$. Funkcia náhodného výberu $\mathrm{T}(X_1, \dots X_n)$ sa nazýva výberová charakteristika (štatistika). Pomocou výberových charakteristík môžeme odhadnúť neznáme charakteristiky náhodných veličín. Dosadením hodnôt $(x_1, \dots x_n)$ realizácie náhodnej veličiny $(X_1, \dots, X_n)$ do štatistiky $T(X_1, \dots X_n)$ dostaneme pozorovanú hodnotu štatistiky $t = T(X_1, \dots X_n)$. [2, 8]

Definujeme preto nasledujúce charakteristiky:

1. Výberový priemer

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1.6}$$

2. Výberový rozptyl

$$S^2(X) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{1.7}$$

3. Výberová smerodajná odchýlka

$$S(X) = \sqrt{S^2(X)} \tag{1.8}$$

4. Výberový koeficient kovariancie

$$S(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})\,(Y_i - \bar{Y}) \tag{1.9}$$

5. Výberový koeficient korelácie

$$R(X,Y) = \frac{S(X,Y)}{S(X)S(Y)}. \tag{1.10}$$

Označme $r(X,Y)$ ako realizáciu výberového korelačného koeficientu $R(X,Y)$. Pre jeho lepšiu slovnú interpretáciu sa niekedy používa nasledujúca stupnica, ktorá klasifikuje mieru vzájomnej lineárnej závislosti. Je však nutné poznamenať, že táto stupnica je iba orientačná, a nie je podložená žiadnym objektívnym kritériom. [7]

$\lvert r(X,Y)\rvert \leq 0{,}3$	nízky stupeň závislosti
$0{,}3 < \lvert r(X,Y)\rvert \leq 0{,}5$	stredný stupeň závislosti
$0{,}5 < \lvert r(X,Y)\rvert \leq 0{,}7$	významný stupeň závislosti
$0{,}7 < \lvert r(X,Y)\rvert \leq 0{,}9$	vysoký stupeň závislosti
$0{,}9 < \lvert r(X,Y)\rvert$	veľmi vysoký stupeň závislosti

1.2 Typy korelačných koeficientov

Ako sme už spomenuli, existuje viacero rôznych typov korelačných koeficientov. V tejto podkapitole si zadefinujeme niektoré z nich.

1.2.1 Pearsonov korelačný koeficient

Pearsonov korelačný koeficient je v súčasnosti najdôležitejším a najpoužívanejším koeficientom korelácie. Jeho výpočet sa realizuje z n párových hodnôt (x_i, y_i). Odhadom jeho teoretickej hodnoty $\rho(X, Y)$ je výberový Pearsonov korelačný koeficient $r(X, Y)$ definovaný vzťahom [6]:

$$r(X,Y) = \frac{s(X,Y)}{s(X)s(Y)}. \tag{1.11}$$

Jedná sa o parametrický korelačný koeficient, teda pre jeho správnu interpretáciu by sme mali predpokladať normalitu skúmaných dát. Dvojrozmerná normalita dát znamená, že obe náhodné veličiny majú spoločné dvojrozmerné normálne rozdelenie, ktorého hustota je definovaná vzťahom:

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho)^2}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}, \tag{1.12}$$

kde μ_X, μ_Y sú stredné hodnoty a σ_X, σ_Y sú smerodajné odchýlky jednorozmerných normálnych rozdelení náhodných veličín X a Y. ρ je ich vzájomný korelačný koeficient. [10]

Pokiaľ je hodnota výberového Pearsonovho korelačného koeficientu $|r(X,Y)| = 1$, ležia všetky body na nejakej priamke. Tento koeficient môžeme použiť len na vyjadrenie sily lineárneho vzťahu medzi veličinami. Nezávislé na sile iných vzájomných vzťahov, použitie tohto koeficientu pre účely vyjadrenia vzájomnej závislosti nie je príliš vhodné. Navyše, aj pri veľmi silnom lineárnom vzťahu nám Pearsonov korelačný koeficient neposkytne úplný popis dát. Pre dokonalejší popis by bolo vhodné poznať rovnicu priamky vyjadrujúcu tvar vzťahu. Korelačný koeficient $r(X,Y)$ je veľmi citlivý na odľahlé hodnoty vo vzorke. Môžeme preto dostať veľmi skreslené hodnoty korelácie, ktoré úplne nezodpovedajú skutočnosti. [6]

1.2.2 Spearmanov korelačný koeficient

Keďže je Pearsonov korelačný koeficient citlivý na odľahlé hodnoty vo vzorke, na určenie závislosti dvoch veličín boli navrhnuté neparametrické poradové korelačné koeficienty. Jedným z nich je aj Spearmanov korelačný koeficient. Tento koeficient využíva princíp poradovej korelácie. Jeho podstata spočíva v nahradení realizácií náhodných veličín x_i, y_i ich poradím R_{x_i}, R_{y_i} vzhľadom k ostatným hodnotám výberu zoradeného podľa veľkosti. Pochopiteľne, môže nastať aj situácia, že sa hodnoty x_i, prípadne y_i, opakujú viackrát. Vtedy im priradíme priemernú hodnotu poradí, ktorú by mali, ak by nasledovali za sebou a boli rôzne. Odhadom teoretickej hodnoty Spearmanovho korelačného koeficientu ρ_s je výberový korelačný koeficient r_s definovaný vzťahom [6]:

$$r_s = \rho(R_x, R_y) = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n(n^2-1)}, \tag{1.13}$$

kde D_i sú rozdiely poradí R_{x_i}, R_{y_i} a n je rozsah výberu.

Spearmanov korelačný koeficient, na rozdiel od Pearsonovho korelačného koeficientu, ktorý popisuje iba lineárne vzťahy, zachytáva monotónne vzťahy. To znamená, že zachytáva vzťahy vo všeobecnosti rastúce, prípadne klesajúce. Koeficient je navyše vďaka princípu poradovej korelácie odolný voči odľahlým hodnotám vo vzorke. Pre Pearsonov a Spearmanov výberový korelačný koeficient platí v prípade dvojrozmerného normálneho rozdelenia nasledujúci približný vzťah [6]:

$$r = 2\,sin\left(\frac{\pi}{6} r_s\right). \tag{1.14}$$

1.2.3 Kendallov korelačný koeficient

Ďalším typom korelačného koeficientu je Kendallov korelačný koeficient. Opäť sa jedná o neparametrický koeficient. Vyjadruje silu závislosti medzi dvoma poradovými premennými. To znamená, že ho môžeme interpretovať ako rozdiel medzi pravdepodobnosťou, že hodnoty dvoch premenných sú v rovnakom poradí, oproti pravdepodobnosti, že hodnoty premenných nie sú v rovnakom poradí. [11]

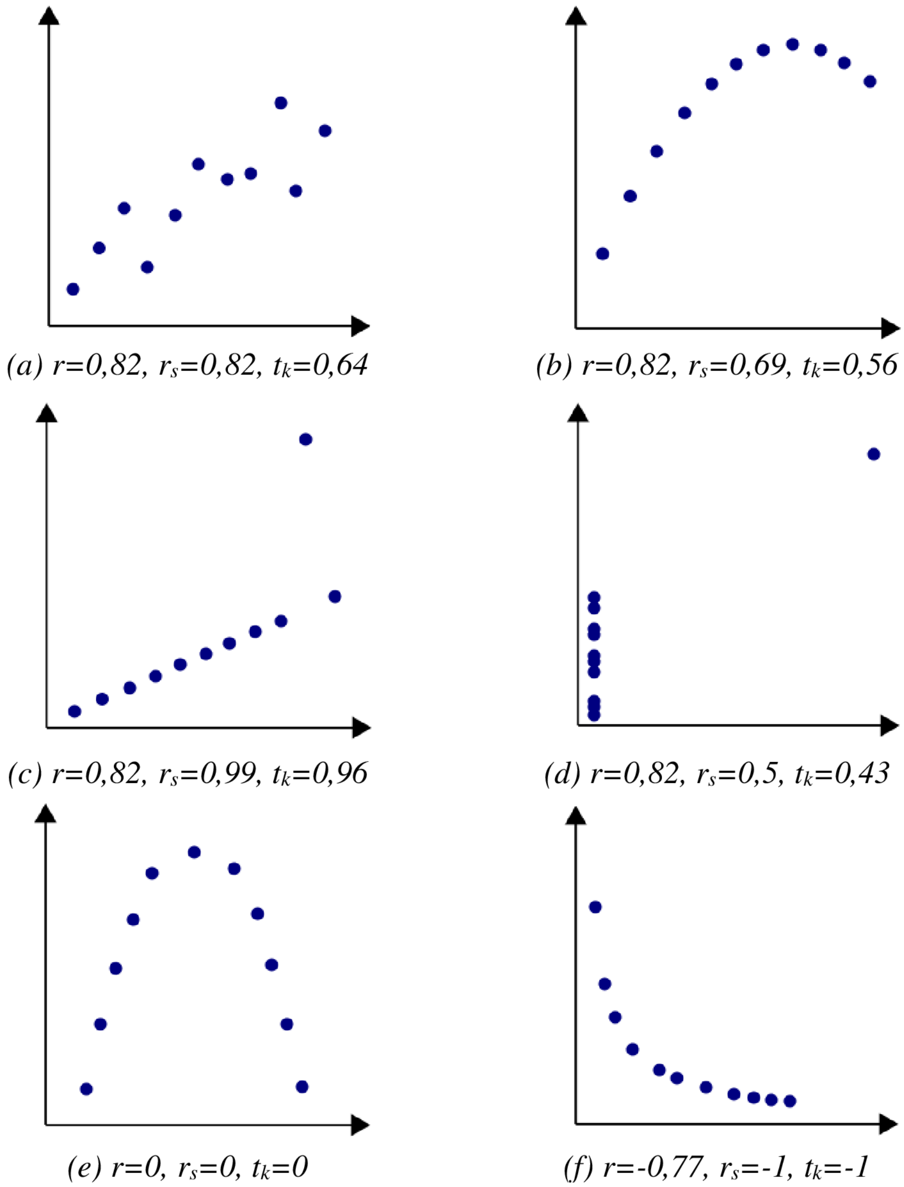
Ako sme už uviedli, podstata Spearmanovho koeficientu spočíva v poradovej korelácii. Kendallov koeficient je založený na inverziách v poradí. Základom je pre nás opäť súbor o rozsahu n hodnôt (x_i, y_i). V prvom kroku zoradíme dvojice (x_i, y_i) tak, aby hodnoty x_i tvorili rastúcu postupnosť. Následne budeme skúmať to, či aj hodnoty y_i majú vzostupnú tendenciu (kladná asociácia medzi veličinami X a Y), prípadne zostupnú tendenciu (záporná asociácia). Kendallov korelačný koeficient teda rozlišuje prípad, kedy $y_j > y_i$, respektíve $y_j < y_i$, pokiaľ $j > i$ $(i = 1,2,\dots,n-1)$. Prvý prípad je nazývaný konkordancia (prípad kladnej asociácie). V druhom prípade nastáva diskordancia (prípad zápornej asociácie). Počet všetkých konkordancií označíme P, počet všetkých diskordancií Q. Rozdiel $S = P - Q$ býva označovaný aj ako Kendallovo S a býva interpretovaný ako jednoduchá miera závislosti. Znamienko S potom prirodzene závisí na tom, či vo vzorke prevládajú konkordancie, prípadne diskordancie. Hodnoty, ktoré môže S nadobúdať, závisia na rozsahu výberu n. Pre potreby výpočtu korelačného koeficientu je nutné vykonať jednoduchú úpravu. Odhadom teoretickej hodnoty koeficientu τ_k je Kendallov koeficient t_k definovaný vzťahom [6]:

$$t_k = \frac{S}{D} = \frac{P - Q}{D}, \tag{1.15}$$

kde D je maximálny možný počet konkordancií, respektíve diskordancií, ktorý je pri rozsahu výberu n rovný hodnote $n(n-1)/2$.

Aj keď výpočet τ_k je pomerne prácny, jeho interpretácia je priamočiarejšia než u Spearmanovho korelačného koeficientu ρ_s. Ak sa $\tau_k = p$, môžeme u dvoch náhodne vybraných jednotiek očakávať s pravdepodobnosťou p, že ich zoradenie podľa kritéria X bude rovnaké ako zoradenie podľa kritéria Y. Poznamenajme, že koeficient τ_k nadobúda vo všeobecnosti menšie absolútne hodnoty než koeficient ρ_s. [6, 11]

Chovanie Pearsonovho, Spearmanovho a Kendallovho korelačného koeficientu pri rôznom rozložení skúmaných dát nám lepšie ilustruje nasledujúci obrázok:



Obrázok 1.3: Zobrazenie rôznych bodových konfigurácií a k ním dopočítaného Pearsonovho (r), Spearmanovho (r_s) a Kendallovho (t_k) korelačného koeficientu [6]

1.2.4 Bodovo biseriálny korelačný koeficient

Biseriálny korelačný koeficient ρ_{pb} vyjadruje vzťah medzi spojitou metrickou premennou a binárnou premennou (nadobúdajúcou iba dve možné hodnoty). Tento koeficient je rezistentný voči miernemu porušeniu normality. Princíp jeho výpočtu spočíva v tom, že n dvojíc merania sa rozdelí na dve skupiny podľa hodnoty alternatívneho parametru. Následne sa spočíta koeficient r_{pb}, ktorý je odhadom teoretickej hodnoty ρ_{pb}, podľa vzťahu [4, 6]:

$$r_{pb} = \frac{(\bar{x}_1 - \bar{x}_2)}{s}\sqrt{\frac{n_1 n_2}{n(n-1)}}, \tag{1.16}$$

kde n_i sú počty spojitého parametru v oboch skupinách, $\bar{x}_i$ sú jeho priemerné hodnoty a s je smerodajná odchýlka súboru. Ďalšie informácie k tejto problematike sa dočítame v [4].

2 Testy hypotéz o korelačných koeficientoch

Neoddeliteľnou súčasťou matematickej štatistiky sú aj testy rôznych štatistických hypotéz. V našom prípade sa budeme zaoberať testami súvisiacimi s korelačnými koeficientami definovanými v predchádzajúcej kapitole. Tieto testy nám v prevažnej miere poskytujú informácie o tom, či môžeme alebo nemôžeme zamietnuť hypotézu o nulovej hodnote korelačného koeficientu (znamenajúcu nezávislosť veličín X a Y).

2.1 Test významnosti korelačného koeficientu

Predpokladajme, že náhodný vektor (X, Y) o rozsahu n prvkov má dvojrozmerné normálne rozdelenie s korelačným koeficientom $\rho(X, Y)$. Pomocou náhodného výberu o n zložkách definujeme výberový korelačný koeficient $R(X, Y)$. S tým súvisí otázka, či získané výsledky môžeme zovšeobecniť a interpretovať pre celý základný súbor, nakoľko sa jedná len o výberový korelačný koeficient. Práve pre tieto účely realizujeme test významnosti korelačného koeficientu. [1, 11]

Testujeme hypotézu

$$H: \rho = 0$$

oproti niektorej z alternatívnych hypotéz:

a) $H_A: \rho \neq 0$

b) $H_A: \rho < 0$

c) $H_A: \rho > 0$.

Testovacia štatistika má pre $n \geq 3$ tvar

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}. \tag{2.1}$$

Táto štatistika má pri platnosti testovanej hypotézy H Studentovo rozdelenie s $(n-2)$ stupňami voľnosti. Hypotézu zamietneme na hladine významnosti α, pokiaľ sa hodnota testovacej štatistiky nachádza v kritickom obore, ktorý má pre jednotlivé prípady a) – c), ktoré sme uviedli vyššie, tvar [1, 11]:

a) $W = (-\infty, -t_{1-\alpha/2}(n-2)\rangle \cup \langle t_{1-\alpha/2}(n-2), \infty)$

b) $W = (-\infty, -t_{1-\alpha}(n-2)\rangle$

c) $W = \langle t_{1-\alpha}(n-2), \infty)$.

Zamietnutie hypotézy H na hladine významnosti α v prospech alternatívnej hypotézy H_A znamená v prípade a) závislosť náhodných veličín X a Y. V prípade b) znamená, že medzi veličinami X a Y existuje záporná korelácia. V prípade c) môžeme prehlásiť na hladine významnosti α, že medzi veličinami X a Y existuje kladná korelácia. [1, 11]

Rovnakým spôsobom môžeme realizovať aj test významnosti bodovo biseriálneho korelačného koeficientu. V testovacej štatistike (2.1) potom nahradíme výberový korelačný koeficient r koeficientom r_{pb}. Štatistika t sa opäť riadi Studentovým rozdelením s $(n-2)$ stupňami voľnosti. [13]

2.2 Test významnosti Spearmanovho korelačného koeficientu

Ako sme už spomenuli v kapitole 1, v prípade, že nemôžeme predpokladať normalitu u veličín X a Y, je vhodnejšie pre účely korelácie použiť Spearmanov korelačný koeficient. Na posúdenie jeho štatistickej významnosti nám poslúži nasledujúci test.

Testujeme hypotézu

$$H: \rho_s = 0$$

oproti niektorej z alternatívnych hypotéz:

a) $H_A: \rho_s \neq 0$

b) $H_A: \rho_s < 0$

c) $H_A: \rho_s > 0.$

Pre rozsahy výberov $n \leq 30$ ako testovaciu štatistiku využijeme výberový korelačný koeficient r_s definovaný vzťahom (1.13). Kritický obor má pre prípady a) – c) nasledujúci tvar:

a) $W = \langle -1, -R_{s,1-\alpha/2} \rangle \cup \langle R_{s,1-\alpha/2}, 1 \rangle$

b) $W = \langle -1, -R_{s,1-\alpha} \rangle$

c) $W = \langle R_{s,1-\alpha}, 1 \rangle,$

kde R_s sú kritické hodnoty, ktoré pre danú hladinu významnosti nájdeme v prílohe A. Hypotézu zamietame na hladine významnosti α, pokiaľ sa hodnota koeficientu r_s nachádza v kritickom obore. [2]

Pre rozsah výberu $n > 20$ je možné taktiež použiť modifikovanú testovaciu štatistiku (2.1), v ktorej nahradíme výberový korelačný koeficient r koeficientom r_s [2]:

$$t = \frac{r_s \sqrt{n-2}}{\sqrt{1 - r_s^2}}. \tag{2.2}$$

V prípade platnosti hypotézy H má táto štatistika asymptoticky Studentovo rozdelenie s $(n-2)$ stupňami voľnosti. Kritický obor, na základe ktorého môžeme rozhodnúť na hladine významnosti α o prípadnom zamietnutí hypotézy H, má pre prípady a) – c) tvar:

a) $W = (-\infty, -t_{1-\alpha/2}(n-2) \rangle \cup \langle t_{1-\alpha/2}(n-2), \infty)$

b) $W = (-\infty, -t_{1-\alpha}(n-2) \rangle$

c) $W = \langle t_{1-\alpha}(n-2), \infty).$

Pre rozsah výberu $n > 30$ môžeme použiť aj testovaciu štatistiku

$$u = r_s \sqrt{n-1}. \tag{2.3}$$

Táto štatistika má pri platnosti skúmanej hypotézy asymptoticky normované normálne rozdelenie. Kritický obor má pre prípady a) – c) tvar [2]:

a) $W = (-\infty, -u_{1-\alpha/2} \rangle \cup \langle u_{1-\alpha/2}, \infty)$

b) $W = (-\infty, -u_{1-\alpha} \rangle$

c) $W = \langle u_{1-\alpha}, \infty).$

2.3 Test významnosti Kendallovho korelačného koeficientu

Rovnako ako ostatné definované korelačné koeficienty, aj Kendallov korelačný koeficient môžeme testovať pre účely posúdenia jeho štatisticky preukázateľnej odlišnosti od nuly. Z tohto dôvodu testujeme hypotézu

$$H: \tau_k = 0$$

oproti niektorej z alternatívnych hypotéz:

a) $H_A: \tau_k \neq 0$

b) $H_A: \tau_k < 0$

c) $H_A: \tau_k > 0.$

Pre malé rozsahy výberov môžeme opäť na rozhodnutie o platnosti testovanej hypotézy na hladine významnosti α využiť tabuľkové kritické hodnoty T_k, ktoré môžeme nájsť v prílohe A. Tvary kritického oboru sú pre prípady a) – c) nasledujúce [11]:

a) $W = \langle -1, -T_{k,1-\alpha/2} \rangle \cup \langle T_{k,1-\alpha/2}, 1 \rangle$

b) $W = \langle -1, -T_{k,1-\alpha} \rangle$

c) $W = \langle T_{k,1-\alpha}, 1 \rangle.$

Hypotézu H zamietneme na hladine významnosti α, pokiaľ hodnota Kendallovho korelačného koeficientu $t_k \in W$.

Rovnako ako v prípade Spearmanovho korelačného koeficientu, aj u Kendallovho korelačného koeficientu môžeme s rastúcim rozsahom výberu n použiť štatistiku podliehajúcu jednému z tabuľkových rozdelení pravdepodobnosti. V tomto prípade platí, že pre väčší rozsah výberu (približne $n > 60$) môžeme použiť štatistiku

$$u = \frac{3t_k\sqrt{n(n-1)}}{\sqrt{2(2n+5)}}, \tag{2.4}$$

ktorá má pri platnosti testovanej hypotézy H asymptoticky normované normálne rozdelenie. Hypotézu H zamietame na hladine významnosti α, pokiaľ $u \in W$. Pre kritický obor W teda platí [11]:

a) $W = (-\infty, -u_{1-\alpha/2} \rangle \cup \langle u_{1-\alpha/2}, \infty)$

b) $W = (-\infty, -u_{1-\alpha} \rangle$

c) $W = \langle u_{1-\alpha}, \infty).$

2.4 Test na porovnanie korelačného koeficientu s danou konštantou

Okrem testov významnosti (nezávislosti) korelačných koeficientov existuje množstvo ďalších testov spojených s korelačnými koeficientami. Jedným z nich je aj test, pri ktorom chceme odhadnúť koreláciu $\rho(X, Y)$ zložiek náhodného vektoru $(\mathbb{X}, \mathbb{Y})$. Predpokladajme, že náhodný vektor má dvojrozmerné normálne rozdelenie. Označme $r(X, Y)$ ako realizáciu výberového korelačného koeficientu. Predpokladajme, že platí: $n \geq 10, |r| \neq 1, |\rho_0| \neq 1$. [8]

Testujeme hypotézu

$$H: \rho = \rho_0$$

oproti niektorej z alternatívnych hypotéz:

a) $H_A: \rho \neq \rho_0$

b) $H_A: \rho < \rho_0$

c) $H_A: \rho > \rho_0$.

Testovacie kritérium je

$$u = \left(ln\left(\frac{1+r}{1-r}\right) - ln\left(\frac{1+\rho_0}{1-\rho_0}\right) - \frac{\rho_0}{n-1} \right) \frac{\sqrt{n-3}}{2}. \tag{2.5}$$

Štatistika u má pri platnosti hypotézy H asymptoticky normované normálne rozdelenie. Hypotézu H zamietneme na hladine významnosti α v prípade, že $u \in W$. Kritický obor je pre jednotlivé prípady a) – c) tvaru [8]:

a) $W = (-\infty, -u_{1-\alpha/2}\rangle \cup \langle u_{1-\alpha/2}, \infty)$

b) $W = (-\infty, -u_{1-\alpha}\rangle$

c) $W = \langle u_{1-\alpha}, \infty)$.

2.5 Test rovnosti dvoch korelačných koeficientov

Nasledujúci test môžeme interpretovať aj ako test rozdielu dvoch korelačných koeficientov. Pomocou tohto testu budeme zisťovať, či oba skúmané nezávislé náhodné výbery pochádzajú z rovnakého základného súboru. Predpokladajme, že tieto náhodné výbery majú dvojrozmerné normálne rozdelenie o rozsahu n_1, respektíve n_2. Označme r_1, respektíve r_2 výberové korelačné koeficienty jednotlivých výberov. [3, 10] Testujeme hypotézu

$$H: \rho_1 = \rho_2$$

oproti niektorej z alternatívnych hypotéz:

a) $H_A: \rho_1 \neq \rho_2$

b) $H_A: \rho_1 < \rho_2$

c) $H_A: \rho_1 > \rho_2$.

Testovacia štatistika je tvaru

$$u = \frac{z_1 - z_2}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}, \tag{2.6}$$

kde

$$z_1 = \frac{1}{2} ln\left(\frac{1+r_1}{1-r_1}\right) \qquad z_2 = \frac{1}{2} ln\left(\frac{1+r_2}{1-r_2}\right).$$

V prípade platnosti hypotézy H má štatistika u asymptoticky normované normálne rozdelenie. Hypotézu H zamietame na hladine významnosti α, ak hodnota testovacej štatistiky u leží v kritickom obore W. Kritický obor pre rozhodnutie o platnosti hypotézy H má pre prípady a) – c) tvar [3]:

a) $W = (-\infty, -u_{1-\alpha/2}\rangle \cup \langle u_{1-\alpha/2}, \infty)$

b) $W = (-\infty, -u_{1-\alpha}\rangle$

c) $W = \langle u_{1-\alpha}, \infty)$.

3 Porovnanie chovania korelačných koeficientov

Využitie korelácie na vyjadrenie závislosti dvoch veličín je v dnešnej dobe naozaj široké, často je využívaná napríklad v medicíne. Problémom je však skutočnosť, že za určitých okolností môžeme dostať skreslené hodnoty korelačného koeficientu, ktoré môžu viesť k nesprávnym záverom ohľadom skúmanej závislosti.

V tejto kapitole pomocou simulácií v programe MATLAB porovnáme chovanie parametrického, čiže Pearsonovho, a neparametrického, Spearmanovho, korelačného koeficientu v závislosti na ich hodnote, rozsahu výberu a ďalších okolnostiach. Vo všetkých prípadoch nám ako základ poslúžia náhodne vygenerované dáta s dvojrozmerným normálnym rozdelením. Na tieto účely sme v MATLABe využili príkaz $mvnrnd(\boldsymbol{\mu}, \boldsymbol{\Sigma}, n)$, kde $\boldsymbol{\mu}$ je vektor stredných hodnôt normálnych náhodných veličín a $\boldsymbol{\Sigma}$ je kovariančná matica, charakterizujúca dvojrozmerné normálne rozdelenie, ktorá má nasledujúci tvar:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}.$$

V našom prípade boli hodnoty nastavené nasledovne: $\mu_X = \mu_Y = 0, \sigma_X^2 = \sigma_Y^2 = 4$. Platilo teda:

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 4 & 4\rho \\ 4\rho & 4 \end{bmatrix}.$$

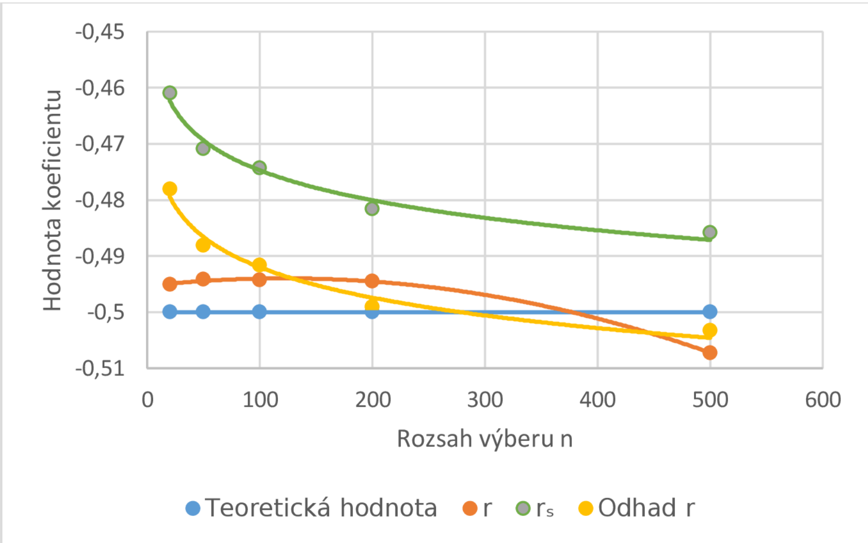
Samotná podstata simulácií spočívala vo vygenerovaní dát s dvojrozmerným normálnym rozdelením, ktoré sme pri danom rozsahu výberu zopakovali 35-krát. Pre každé jedno vygenerovanie bol spočítaný výberový Pearsonov a Spearmanov korelačný koeficient. Zároveň bola v každom generovaní overená normalita dvojrozmerných dát. Na to sme využili dvojvýberový Kolmogorovov-Smirnovov test na hladine významnosti $\alpha = 0{,}05$ a príkaz $kstest2(X, Y)$. Bližšie informácie o tomto teste môžeme nájsť napríklad v [5, str. 211]. Následne sme overili, či spočítané hodnoty výberového Pearsonovho a Spearmanovho korelačného koeficientu majú normálne rozdelenie. K tomu sme využili jednovýberový Kolmogorovov-Smirnovov test na hladine významnosti $\alpha = 0{,}05$ a príkaz $kstest(X)$. Dôvodom pre toto overenie bola skutočnosť, že v prípade normálneho rozdelenia je výberový priemer zo spočítaných korelačných koeficientov bodovým odhadom Pearsonovho a Spearmanovho korelačného koeficientu. Okrem toho sme pomocou vzťahu (1.14) odhadli hodnotu výberového Pearsonovho koeficientu pomocou Spearmanovho. Celý tento postup sme zopakovali pri rôznych hodnotách rozsahu výberu n a rôznych hodnotách ρ.

3.1 Porovnanie korelačných koeficientov pri zachovaní normality

Na úvod sa pozrieme na najjednoduchší prípad, kedy je zachovaná dvojrozmerná normalita skúmaných dát. Rozsahy výberu n boli postupne $n = 20; 50; 100; 200; 500$. Teoretické hodnoty korelačného koeficientu ρ v kovariančnej matici $\boldsymbol{\Sigma}$ sme volili postupne nasledovne: $\rho = -0{,}9; -0{,}5; -0{,}1; 0{,}2; 0{,}4; 0{,}8$. Nasledujúci obrázok nám zobrazuje porovnanie Pearsonovho r, Spearmanovho r_s a odhadu Pearsonovho koeficientu r s teoretickou hodnotou $\rho = -0{,}5$. Pre lepšiu ilustráciu sme spočítané hodnoty približne preložili spojnicou.

Tabuľka 3.1: Hodnoty koeficientov pre $\rho = -0{,}5$ pri zachovaní normality

Rozsah výberu n	**20**	**50**	**100**	**200**	**500**
Teoretická hodnota ρ	-0,5000	-0,5000	-0,5000	-0,5000	-0,5000
r	-0,4951	-0,4941	-0,4942	-0,4945	-0,5072
r_s	-0,4610	-0,4709	-0,4743	-0,4816	-0,4858
Odhad r	-0,4781	-0,4881	-0,4916	-0,4990	-0,5032



Obrázok 3.1: Porovnanie koeficientov pre $\rho = -0{,}5$ pri zachovaní normality

Ako je z obrázku 3.1 zrejmé, pri zachovaní normality sa Pearsonov koeficient r naozaj presne približuje teoretickej hodnote ρ. Spearmanov koeficient r_s sa s rastúcim rozsahom výberu n spresňuje, avšak ani pri rozsahu $n > 100$ sa presnosťou nevyrovná Pearsonovmu koeficientu. Oba koeficienty nadobúdajú menšie absolútne hodnoty ako je teoretická hodnota ρ. Čo sa týka odhadu koeficientu r podľa vzťahu (1.14), pri rozsahoch výberu $n > 100$ sa jeho hodnota blíži spočítanej hodnote r.

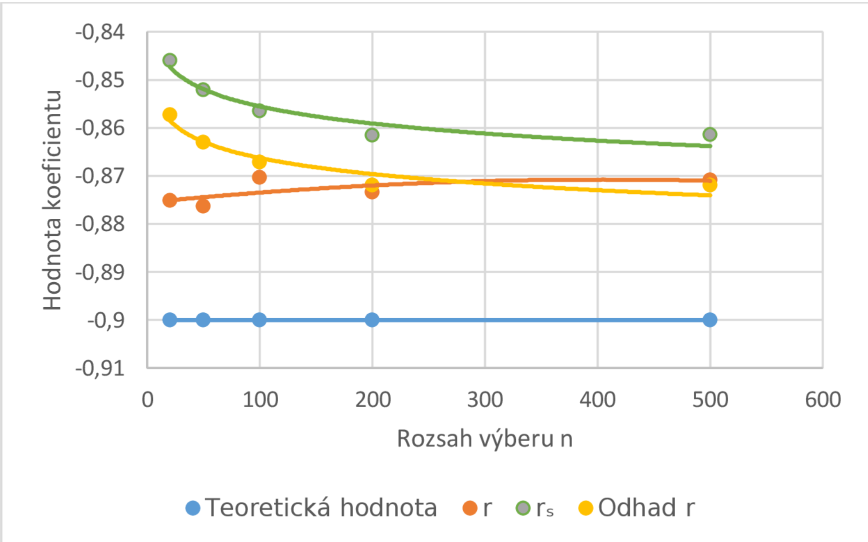
Pri ďalších teoretických hodnotách ρ bolo chovanie uvažovaných koeficientov podobné, Pearsonov koeficient poskytoval naozaj presný popis vzťahu medzi veličinami, zatiaľ čo Spearmanov koeficient sa s rastúcim rozsahom výberu spresňoval. Tieto skutočnosti môžeme vidieť na grafoch a v tabuľkách v prílohe A.

3.2 Porovnanie korelačných koeficientov pri porušení normality

Keďže medzi predpoklady použitia Pearsonovho korelačného koeficientu patrí dvojrozmerná normalita skúmaných dát, pozrieme sa na prípad, kedy je táto normalita porušená. Rozsahy výberu boli opäť $n = 20; 50; 100; 200; 500$. Teoretické hodnoty koeficientu ρ sme volili postupne $\rho = -0{,}9; -0{,}5; -0{,}1; 0{,}2; 0{,}4; 0{,}8$. Najprv sme dosiahli porušenie dvojrozmernej normality pripočítaním náhodne vygenerovaných jednorozmerných dát s odlišným rozdelením ako je normálne rozdelenie k jednej zo súradníc vygenerovaných dvojrozmerných dát. V našom prípade sme pripočítali k súradnici x náhodne vygenerované jednorozmerné dáta s Weibullovým rozdelením. Podrobnejšie informácie o Weibullovom rozdelení môžeme nájsť napríklad v [11, str. 94]. Na tento účel sme využili v MATLABe príkaz $wblrnd(\lambda, k, n)$. Konkrétne sme parametre λ a k zvolili nasledovne: $\lambda = 4, k = 8$. Chovanie koeficientov môžeme pri hodnote $\rho = -0{,}9$ porovnať z nasledujúcej tabuľky a obrázku 3.2.

Tabuľka 3.2: Hodnoty koeficientov pre $\rho = -0{,}9$ pri porušení normality jednou zložkou

Rozsah výberu n	**20**	**50**	**100**	**200**	**500**
Teoretická hodnota ρ	-0,9000	-0,9000	-0,9000	-0,9000	-0,9000
r	-0,8751	-0,8763	-0,8703	-0,8734	-0,8709
r_s	-0,8460	-0,8521	-0,8565	-0,8616	-0,8615
Odhad r	-0,8573	-0,8630	-0,8671	-0,8720	-0,8719



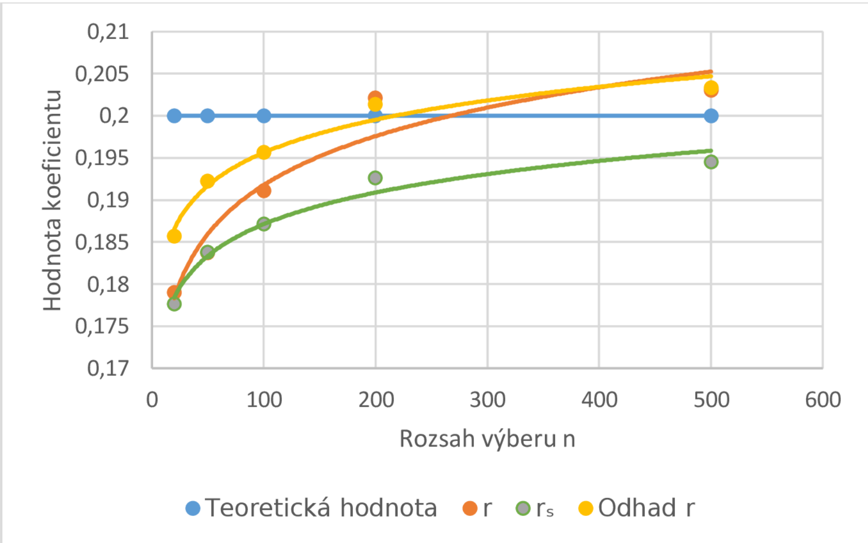
Obrázok 3.2: Porovnanie koeficientov pre $\rho = -0{,}9$ pri porušení normality jednou zložkou

Z obrázku 3.2 môžeme usúdiť, že za podmienok, pri ktorých bola simulácia realizovaná, je Pearsonov koeficient presnejší aj napriek faktu, že pripočítaním dát s Weibullovým rozdelením k zložke x bola podľa výsledkov testu Kolmogorov-Smirnov dvojrozmerná normalita výrazne porušená. Spearmanov koeficient, ktorý naopak dvojrozmernú normalitu dát nepožaduje, sa síce s rastúcim rozsahom výberu spresňoval, avšak stále poskytoval menej presný popis teoretickej hodnoty korelácie.

Mierne odlišnú situáciu môžeme vidieť pri nižšej hodnote koeficientu ρ. Na nasledujúcom obrázku vidíme porovnanie koeficientov pri $\rho = 0{,}2$. Z tabuľky 3.3 môžeme pozorovať, že s nižšou hodnotou ρ sa Spearmanov koeficient presnosťou blíži Pearsonovmu. Taktiež je možné vyčítať, že s rastúcim rozsahom výberu oba koeficienty r a r_s naozaj presne popisujú teoretickú hodnotu korelácie aj napriek porušeniu dvojrozmernej normality.

Tabuľka 3.3: Hodnoty koeficientov pre $\rho = 0{,}2$ pri porušení normality jednou zložkou

Rozsah výberu n	**20**	**50**	**100**	**200**	**500**
Teoretická hodnota ρ	0,2000	0,2000	0,2000	0,2000	0,2000
r	0,1790	0,1837	0,1911	0,2021	0,2030
r_s	0,1776	0,1838	0,1871	0,1926	0,1945
Odhad r	0,1857	0,1922	0,1956	0,2013	0,2033



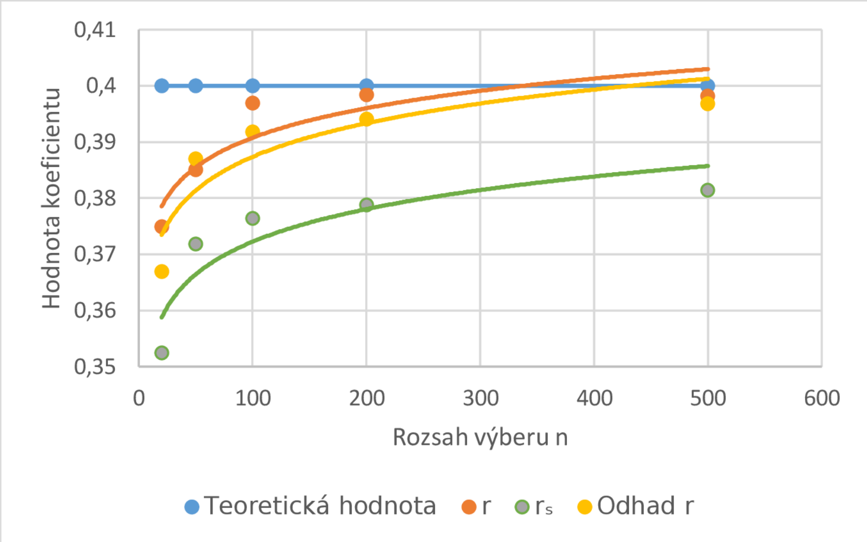
Obrázok 3.3: Porovnanie koeficientov pre $\rho = 0{,}2$ pri porušení normality jednou zložkou

Tabuľky a grafy s porovnaním koeficientov pre zvyšné hodnoty korelácie môžeme nájsť v prílohe A.

Ďalej sa zameriame na prípad, kedy je dvojrozmerná normalita porušená oboma zložkami. To znamená, že k obom súradniciam x a y vygenerovaných dát s dvojrozmerným normálnym rozdelením sme pripočítali náhodne vygenerované jednorozmerné dáta s Weibullovým rozdelením. V zložke x sme zvolili parametre $\lambda = 4$ a $k = 8$. V zložke y boli zvolené parametre $\lambda = 0{,}1$ a $k = 8$. Na nasledujúcom obrázku 3.4 je zobrazený graf pre porovnanie koeficientov pri $\rho = 0{,}4$. Opäť nastáva situácia, kedy je Pearsonov koeficient aj napriek ešte výraznejšiemu porušeniu normality presnejší než Spearmanov, ktorý sa opäť s rastúcim rozsahom výberu spresňuje.

Tabuľka 3.4: Hodnoty koeficientov pre $\rho = 0{,}4$ pri porušení normality oboma zložkami

Rozsah výberu n	**20**	**50**	**100**	**200**	**500**
Teoretická hodnota ρ	0,4000	0,4000	0,4000	0,4000	0,4000
r	0,3750	0,3851	0,3970	0,3985	0,3982
r_s	0,3525	0,3719	0,3765	0,3788	0,3815
Odhad r	0,3670	0,3870	0,3918	0,3941	0,3969

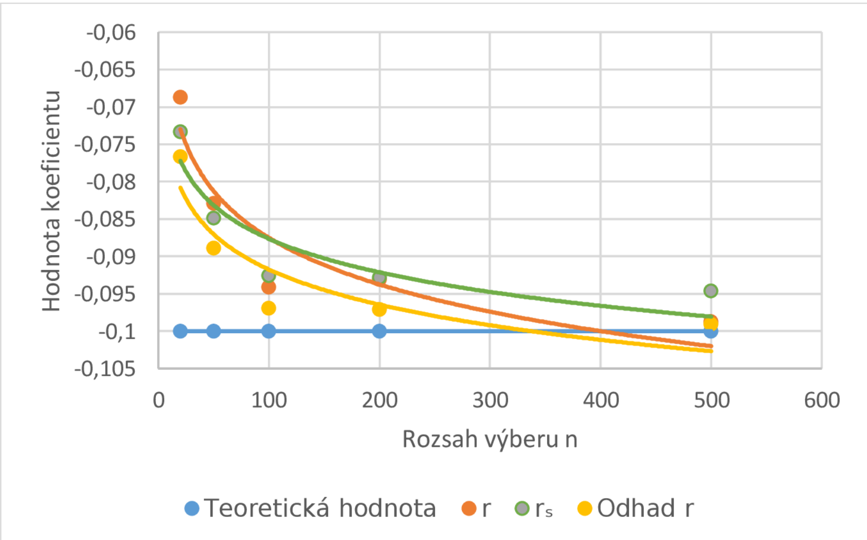


Obrázok 3.4: Porovnanie koeficientov pre $\rho = 0{,}4$ pri porušení normality oboma zložkami

Avšak, rovnako ako v prípade porušenia normality jednou zložkou, aj pri porušení oboma zložkami sa pri klesajúcej absolútnej hodnote korelácie Spearmanov koeficient vyrovnáva Pearsonovmu a poskytuje v podstate rovnako presný popis korelácie. Pri rastúcom rozsahu výberu sa navyše oba koeficienty skutočne presne približujú teoretickej hodnote ρ. Tieto fakty si môžeme pre hodnotu $\rho = -0{,}1$ všimnúť na obrázku 3.5.

Tabuľka 3.5: Hodnoty koeficientov pre $\rho = -0{,}1$ pri porušení normality oboma zložkami

Rozsah výberu n	**20**	**50**	**100**	**200**	**500**
Teoretická hodnota ρ	-0,1000	-0,1000	-0,1000	-0,1000	-0,1000
r	-0,0687	-0,0829	-0,0941	-0,0931	-0,0987
r_s	-0,0733	-0,0849	-0,0926	-0,0928	-0,0946
Odhad r	-0,0767	-0,0889	-0,0969	-0,0971	-0,0991



Obrázok 3.5: Porovnanie koeficientov pre $\rho = -0{,}1$ pri porušení normality oboma zložkami

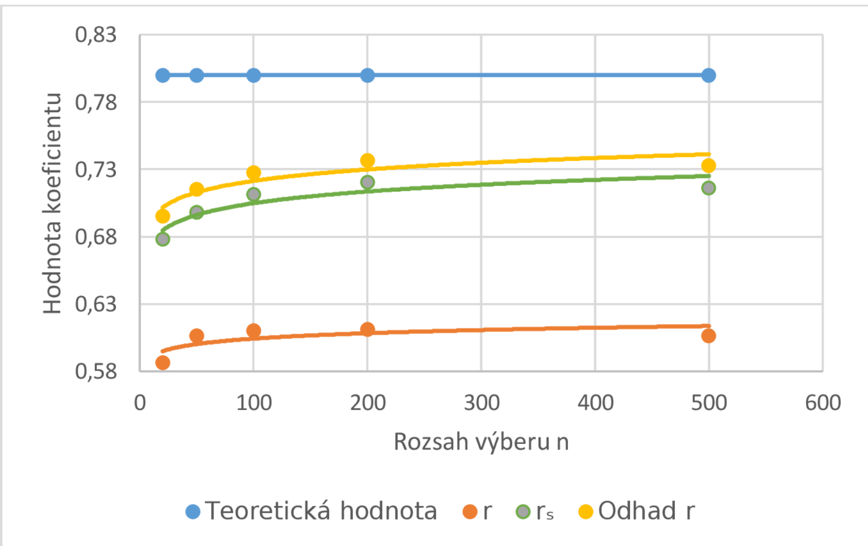
Výsledky pri ostatných hodnotách koeficientov zodpovedali popisu prezentovanému v tejto podkapitole. Grafy a tabuľky pre zvyšné hodnoty ρ nájdeme v prílohe A.

3.3 Porovnanie korelačných koeficientov pri výskyte odľahlých hodnôt

Často býva prezentovaný fakt, že Pearsonov koeficient je citlivý na odľahlé hodnoty, pričom Spearmanov koeficient je voči odľahlým hodnotám rezistentný. V tejto podkapitole teda porovnáme tieto dva koeficienty pri výskyte odľahlých hodnôt vo vzorke. Za základ nám poslúžili dáta s dvojrozmerným normálnym rozdelením s parametrami definovanými v úvode tejto kapitoly. Hodnoty koeficientu ρ boli postupne $\rho = -0{,}9; -0{,}5; -0{,}1; 0{,}2; 0{,}4; 0{,}8$. Rozsahy výberu sme postupne volili $n = 20; 60; 100; 200; 500$. Pri každom koeficiente sme vždy pri danom rozsahu pripočítali k 5 % bodov k súradnici y hodnotu 8. Na obrázku 3.6 vidíme situáciu, kedy je $\rho = 0{,}8$. Ako si môžeme povšimnúť, pri výskyte odľahlých hodnôt v skúmanej vzorke je Spearmanov koeficient skutočne presnejším popisom teoretickej hodnoty korelácie. Avšak aj pri ňom sa vyskytujú pomerne značné odchýlky od hodnoty ρ. Z dôvodu väčšej presnosti Spearmanovho koeficientu sa taktiež odhad Pearsonovho koeficientu pomerne dosť líši od jeho skutočne vypočítanej hodnoty.

Tabuľka 3.6: Hodnoty koeficientov pre $\rho = 0{,}8$ pri výskyte odľahlých hodnôt

Rozsah výberu n	**20**	**60**	**100**	**200**	**500**
Teoretická hodnota ρ	0,8000	0,8000	0,8000	0,8000	0,8000
r	0,5864	0,6066	0,6104	0,6113	0,6066
r_s	0,6780	0,6982	0,7112	0,7203	0,7164
Odhad r	0,6952	0,7150	0,7276	0,7365	0,7328

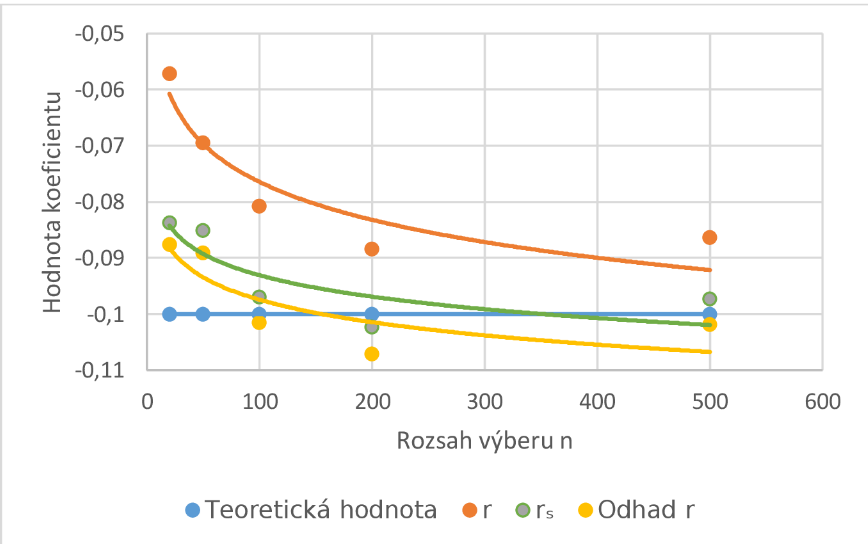


Obrázok 3.6: Porovnanie koeficientov pre $\rho = 0{,}8$ pri výskyte odľahlých hodnôt

Pri nižších hodnotách ρ je nepresnosť Spearmanovho koeficientu výrazne nižšia, s rastúcim rozsahom výberu už naozaj presne popisuje teoretickú hodnotu. Pearsonov koeficient má taktiež nižšiu odchýlku od teoretickej hodnoty, avšak stále zostáva menej presnou voľbou pre popis korelácie. Z toho vyplýva, že pri znalosti faktu o výskyte odľahlých hodnôt v súbore je vhodnejšou voľbou použiť Spearmanov koeficient. Na obrázku 3.7 vidíme porovnanie koeficientov pri $\rho = -0{,}1$. Tabuľky a grafy pre zostávajúce hodnoty nájdeme v prílohe A.

Tabuľka 3.7: Hodnoty koeficientov pre $\rho = -0{,}1$ pri výskyte odľahlých hodnôt

Rozsah výberu n	**20**	**60**	**100**	**200**	**500**
Teoretická hodnota ρ	-0,1000	-0,1000	-0,1000	-0,1000	-0,1000
r	-0,0572	-0,0695	-0,0808	-0,0884	-0,0863
r_s	-0,0837	-0,0851	-0,0970	-0,1023	-0,0973
Odhad r	-0,0876	-0,0891	-0,1015	-0,1071	-0,1019



Obrázok 3.7: Porovnanie koeficientov pre $\rho = -0{,}1$ pri výskyte odľahlých hodnôt

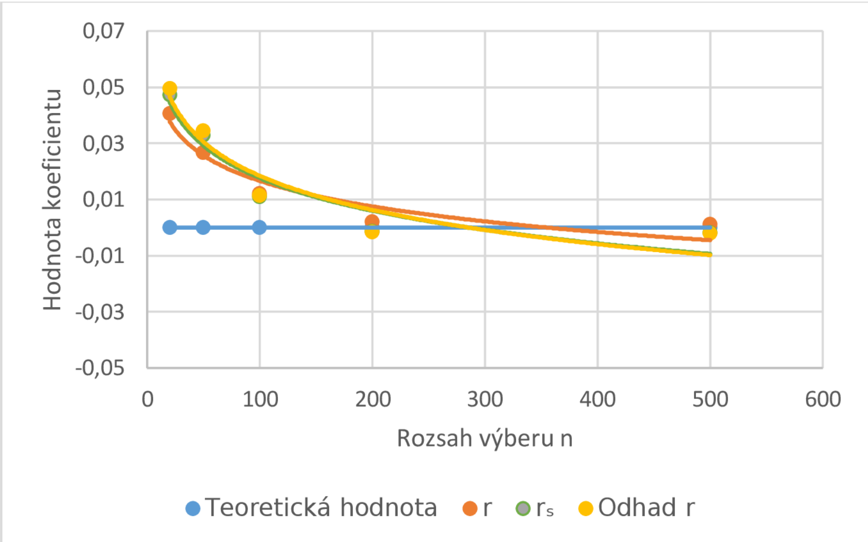
3.4 Porovnanie korelačných koeficientov pri skladaní rozdelení

Na záver sa pozrieme na prípad, kedy je náš pozorovaný súbor tvorený zložením viacerých dát s dvojrozmerným normálnym rozdelením. Podrobnejšie informácie o tejto problematike môžeme nájsť v [9]. My sme sa obmedzili na prípad, kedy sčítame dva súbory s dvojrozmerným normálnym rozdelením. Využili sme pritom váhové funkcie w_i, pri ktorých platí $\sum_{i=1}^{m} w_i = 1$. Výsledné súradnice bodov (x, y) sme teda dostali nasledovným spôsobom: $\mathrm{x} = w_1 x_1 + w_2 x_2$, $y = w_1 y_1 + w_2 y_2$, kde (x_1, y_1) sú súradnice bodov s dvojrozmerným normálnym rozdelením s parametrami $\boldsymbol{\mu}$ a $\boldsymbol{\Sigma}$ definovanými v úvode kapitoly 3 a korelačným koeficientom ρ_1. Obdobne (x_2, y_2) sú súradnice bodov s dvojrozmerným normálnym rozdelením s parametrami $\boldsymbol{\mu}$ a $\boldsymbol{\Sigma}$ definovanými v úvode kapitoly 3 a korelačným koeficientom ρ_2. Výsledný teoretický korelačný koeficient ρ by potom mal nadobúdať hodnotu $\rho = w_1\rho_1 + w_2\rho_2$. Rozsahy výberu boli postupne $n = 20; 50; 100; 200; 500$. Hodnoty váhových funkcií w_i a koeficientov ρ_i budú vždy uvedené samostatne pri každom grafe a tabuľke.

Na obrázku 3.8 je zobrazená situácia, kedy by výsledný koeficient ρ mal byť rovný nule. Ako je vidieť, oba koeficienty r a r_s nadobúdajú prakticky identické hodnoty a veľmi presne popisujú ideálnu hodnotu ρ.

Tabuľka 3.8: Hodnoty koeficientov pre $\rho_1 = 0{,}5;\ \rho_2 = -0{,}5;\ w_1 = 0{,}5;\ w_2 = 0{,}5$

Rozsah výberu n	**20**	**50**	**100**	**200**	**500**
Teoretická hodnota ρ	0,0000	0,0000	0,0000	0,0000	0,0000
r	0,0407	0,0267	0,0122	0,0021	0,0013
r_s	0,0473	0,0330	0,0109	-0,0014	-0,0018
Odhad r	0,0495	0,0345	0,0115	-0,0015	-0,0019

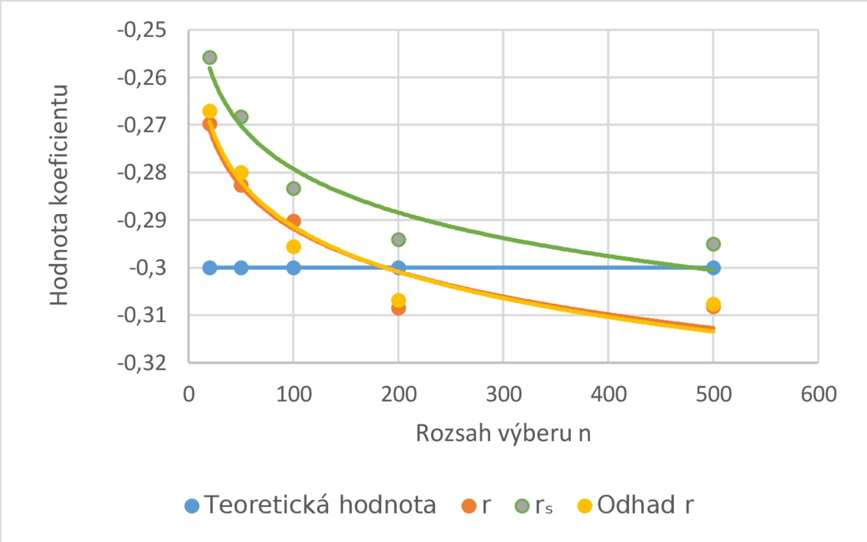


Obrázok 3.8: Porovnanie koeficientov pre $\rho_1 = 0{,}5;\ \rho_2 = -0{,}5;\ w_1 = 0{,}5;\ w_2 = 0{,}5$

Zvolením váhových funkcií $w_1 = w_2 = 0{,}5$ dosiahneme, že Pearsonov a Spearmanov koeficient sa skoro vôbec nebudú líšiť od uvažovanej hodnoty. To nám ilustruje aj nasledujúci obrázok 3.9, kedy uvažujeme hodnotu $\rho = -0{,}3$.

Tabuľka 3.9: Hodnoty koeficientov pre $\rho_1 = 0{,}2;\ \rho_2 = -0{,}8;\ w_1 = 0{,}5;\ w_2 = 0{,}5$

Rozsah výberu n	**20**	**50**	**100**	**200**	**500**
Teoretická hodnota ρ	-0,3000	-0,3000	-0,3000	-0,3000	-0,3000
r	-0,2698	-0,2827	-0,2902	-0,3084	-0,3082
r_s	-0,2558	-0,2682	-0,2833	-0,2941	-0,2950
Odhad r	-0,2670	-0,2799	-0,2955	-0,3068	-0,3077

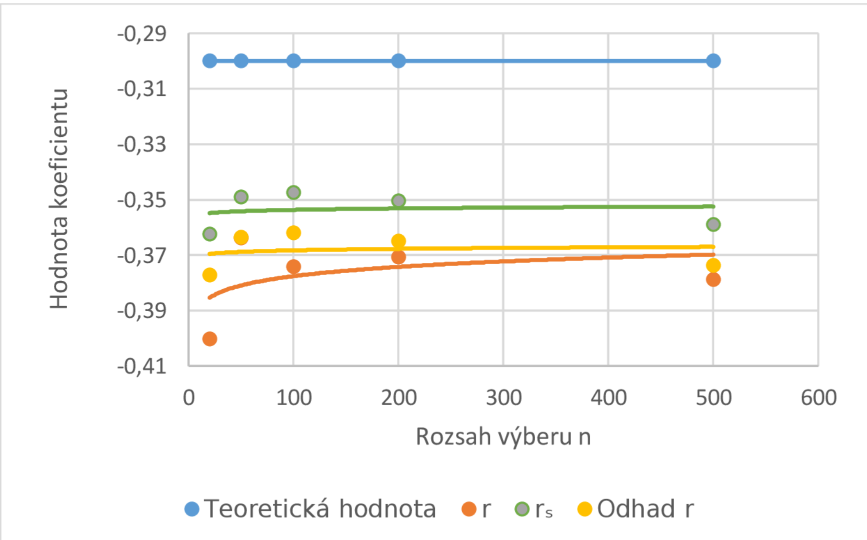


Obrázok 3.9: Porovnanie koeficientov pre $\rho_1 = 0{,}2$; $\rho_2 = -0{,}8$; $w_1 = 0{,}5$; $w_2 = 0{,}5$

Odlišná situácia ale nastáva v prípade, že nezvolíme váhové funkcie rovnomerne na obe strany. V takomto prípade sa už koeficienty r a r_s začínajú výraznejšie odchyľovať od ideálnej hodnoty. O niečo presnejší popis poskytuje Spearmanov koeficient r_s, čo nám ilustruje nasledujúci obrázok 3.10, kedy by opäť teoretická hodnota ρ mala byť $-0{,}3$. Grafy a tabuľky pre zvyšné hodnoty nájdeme v prílohe A.

Tabuľka 3.10: Hodnoty koeficientov pre $\rho_1 = 0{,}6$; $\rho_2 = -0{,}4$; $w_1 = 0{,}1$; $w_2 = 0{,}9$

Rozsah výberu n	**20**	**50**	**100**	**200**	**500**
Teoretická hodnota ρ	-0,3000	-0,3000	-0,3000	-0,3000	-0,3000
r	-0,4003	-0,3639	-0,3741	-0,3708	-0,3789
r_s	-0,3624	-0,3491	-0,3475	-0,3505	-0,3590
Odhad r	-0,3773	-0,3636	-0,3619	-0,3650	-0,3737



Obrázok 3.10: Porovnanie koeficientov pre $\rho_1 = 0{,}6$; $\rho_2 = -0{,}4$; $w_1 = 0{,}1$; $w_2 = 0{,}9$

Záver

Cieľom tejto práce bolo porovnať chovanie parametrických a neparametrických korelačných koeficientov pri rôznom rozložení dát. V úvode sme vysvetlili základné pojmy, ktoré sa týkajú jednoduchej korelačnej analýzy, dôležitého prostriedku na skúmanie závislosti dvoch veličín. Praktickú časť tvorilo práve spomínané porovnanie, ktoré sme realizovali v programe MATLAB na náhodne vygenerovaných dátach s dvojrozmerným normálnym rozdelením a korelačným koeficientom ρ. Ako reprezentanta parametrického korelačného koeficientu sme prirodzene zvolili výberový Pearsonov korelačný koeficient r. V prípade neparametrického korelačného koeficientu sme zvolili výberový Spearmanov korelačný koeficient r_s. Spočítané hodnoty týchto koeficientov sme porovnávali s teoretickou hodnotou ρ pri rôznom rozsahu výberu n a rôznej hodnote ρ.

Praktickú časť sme rozdelili na štyri rôzne prípady. V prvom prípade sme sa zamerali na chovanie koeficientov pri splnení predpokladu dvojrozmernej normality skúmaných dát. Splnenie tohto predpokladu znamenalo väčšiu presnosť u Pearsonovho koeficientu r, ktorý dosahoval vysokú presnosť už pri malých rozsahoch výberu a s rastúcim n sa ešte spresňoval. Spearmanov koeficient ani pri rozsahoch výberu $n > 100$ nedosahoval presnosť spomínaného koeficientu r. Oba tieto koeficienty vo všeobecnosti nadobúdali menšie absolútne hodnoty než hodnota ρ.

V druhom prípade sme predpoklad dvojrozmernej normality porušili pomocou dát s Weibullovým rozdelením. Chovanie oboch koeficientov bolo podobné v prípade porušenia normality jednou zložkou a rovnako aj v prípade porušenia normality oboma zložkami. Vyššia presnosť bola opäť dosiahnutá u Pearsonovho koeficientu, zatiaľ čo koeficient r_s sa viac líšil od teoretickej hodnoty. V prípade nižších hodnôt korelácie, kedy sa hodnota ρ pohybovala okolo 0, sa však tento rozdiel strácal a oba koeficienty sa chovali približne rovnako a pri vyšších rozsahoch výberu už naozaj presne popisovali hodnotu ρ. Avšak, akonáhle sa $|\rho|$ blížila k hodnote 1, koeficienty r a r_s sa markantnejšie líšili od teoretickej hodnoty ρ.

Ďalším prípadom bola situácia, kedy sa vo vzorke vyskytovali odľahlé hodnoty. V tejto situácii bol už Spearmanov koeficient výrazne presnejším popisom hodnoty korelácie. Vysokú presnosť dosahoval najmä pri väčších rozsahoch $n > 100$ za podmienok, kedy sa hodnota ρ pohybovala okolo 0. Pearsonov koeficient sa najmä pri $|\rho|$ blížiacej sa k hodnote 1 výrazne odlišoval od teoretickej hodnoty korelácie.

Ako posledné sme sa pozreli na prípad, kedy bol výber tvorený dvoma súbormi s dvojrozmerným normálnym rozdelením s rôznymi koeficientami ρ_1 a ρ_2. Výsledný teoretický koeficient sme dostali pomocou váhových funkcií w_i. Ako sa ukázalo, tento spôsob popisuje teoretickú hodnotu najmä v prípade, kedy váhové funkcie zvolíme rovnomerne na obe strany súborov s koeficientami ρ_1 a ρ_2. Presnosť koeficientov r a r_s bola v tomto prípade prakticky identická. V prípade nerovnomerného zvolenia váhových funkcií sa už hodnoty r a r_s viac odlišujú od teoretickej hodnoty v prospech väčšej z absolútnych hodnôt súčinu $w_i\rho_i$. O niečo presnejší bol v tejto situácii Spearmanov koeficient r_s.

Simulácie chovania korelačných koeficientov nám ukázali, že v prípade, kedy je splnená dvojrozmerná normalita skúmaných dát je naozaj výhodnejšie použitie Pearsonovho koeficientu, ktorý naozaj presne popisuje teoretickú hodnotu korelácie. Porušenie normality však nemusí automaticky znamenať jeho nižšiu presnosť a voľbu použitia Spearmanovho koeficientu. Je potrebné skúmať, akým spôsobom je normalita porušená. Spearmanov koeficient je však naozaj vhodnejší v prípadoch, kedy sa vo vzorke vyskytujú odľahlé hodnoty, na ktoré, ako sa ukázalo, je Pearsonov koeficient naozaj citlivý.

Zoznam použitých zdrojov

[1] ANDĚL, Jiří. *Základy matematické statistiky*. Vyd. 3. Praha: Matfyzpress, 2011. ISBN 978-80-7378-162-0.

[2] BUDÍKOVÁ, Marie, Maria KRÁLOVÁ a Bohumil MAROŠ. *Průvodce základními statistickými metodami*. Praha: Grada, 2010. Expert (Grada). ISBN 978-80-247-3243-5.

[3] BUDÍKOVÁ, Marie, Tomáš LERCH a Štěpán MIKOLÁŠ. *Základní statistické metody*. Brno: Masarykova univerzita v Brně, 2005. ISBN 80-210-3886-1.

[4] CHENG, Ying a Haiyan LIU. A short note on the maximal point-biserial correlation under non-normality. *British Journal of Mathematical and Statistical Psychology* [online]. 2016, **69**(3), 344-351 [cit. 2018-03-27]. DOI: 10.1111/bmsp.12075. ISSN 00071102. Dostupné z: http://doi.wiley.com/10.1111/bmsp.12075

[5] HEBÁK, Petr. *Statistické myšlení a nástroje analýzy dat*. Praha: Informatorium, 2013. ISBN 978-80-7333-105-4.

[6] HENDL, Jan. *Přehled statistických metod: analýza a metaanalýza dat*. Páté, rozšířené vydání. Praha: Portál, 2015. ISBN 978-80-262-0981-2.

[7] HOLICKÝ, Milan. *Aplikace teorie pravděpodobnosti a matematické statistiky*. V Praze: České vysoké učení technické, 2015. ISBN 978-80-01-05803-9.

[8] KARPÍŠEK, Zdeněk. *Matematika IV*. 4., přeprac. vyd. Brno: Akademické nakladatelství CERM, 2014. ISBN 978-80-214-4858-2.

[9] KOWALSKI, Charles J. On the Effects of Non-Normality on the Distribution of the Sample Product-Moment Correlation Coefficient. *Applied Statistics* [online]. 1972, **21**(1), 1- [cit. 2018-03-27]. DOI: 10.2307/2346598. ISSN 00359254. Dostupné z: http://www.jstor.org/stable/10.2307/2346598?origin=crossref

[10] KOŽÍŠEK, Jan. *Statistická analýza*. Vyd. 4. přeprac. Praha: Vydavatelství ČVUT, 2002. ISBN 80-01-02496-2.

[11] MARKECHOVÁ, Diana, Beata STEHLÍKOVÁ a Anna TIRPÁKOVÁ. *Štatistické metódy a ich aplikácie*. Nitra: UKF, 2011. ISBN 978-80-809-4807-8. Dostupné z: http://www.km.fpv.ukf.sk/upload_publikacie/20120125_143707__1.pdf

[12] Real Statistics. *Real Statistics Using Excel*. [online]. c2014-2017 [cit. 2018-04-24]. Dostupné z: http://www.real-statistics.com/statistics-tables/

[13] TATE, Robert F. Correlation Between a Discrete and a Continuous Variable. Point-Biserial Correlation. *The Annals of Mathematical Statistics* [online]. 1954, **25**(3), 603-607 [cit. 2018-03-27]. DOI: 10.1214/aoms/1177728730. ISSN 0003-4851. Dostupné z: http://projecteuclid.org/euclid.aoms/1177728730

Zoznam príloh

A Obrazová a tabuľková príloha

A.1 Tabuľkové kritické hodnoty pre testy významnosti koeficientov ρ_s a τ_k

Tabuľka A.1: Kritické hodnoty pre testovanie významnosti Spearmanovho korelačného koeficientu [12]

n	$\alpha = 0,05$	$\alpha = 0,01$
5	1,000	–
6	0,886	1,000
7	0,786	0,929
8	0,738	0,881
9	0,700	0,833
10	0,648	0,794
11	0,618	0,755
12	0,587	0,727
13	0,560	0,703
14	0,538	0,679
15	0,521	0,654
16	0,503	0,635
17	0,488	0,618
18	0,472	0,600
19	0,460	0,584
20	0,447	0,570
21	0,436	0,556
22	0,425	0,544
23	0,416	0,532
24	0,407	0,521
25	0,398	0,511
26	0,390	0,501
27	0,383	0,492
28	0,375	0,483
29	0,368	0,475
30	0,362	0,467

Tabuľka A.2: Kritické hodnoty pre testovanie významnosti Kendallovho korelačného koeficientu [12]

$\boldsymbol{n}$	$\boldsymbol{\alpha = 0,05}$	$\boldsymbol{\alpha = 0,01}$
5	1,000	–
6	0,867	1,000
7	0,714	0,905
8	0,643	0,786
9	0,556	0,722
10	0,551	0,644
11	0,491	0,600
12	0,455	0,576
13	0,436	0,564
14	0,407	0,516
15	0,390	0,505
16	0,383	0,483
17	0,368	0,471
18	0,346	0,451
19	0,333	0,439
20	0,326	0,421
21	0,314	0,410
22	0,307	0,394
23	0,296	0,391
24	0,290	0,377
25	0,287	0,367
26	0,280	0,360
27	0,271	0,356
28	0,265	0,344
29	0,261	0,340
30	0,255	0,333

A.2 Tabuľky pre zvyšné simulácie

Tabuľka A.3: Hodnoty koeficientov pri zachovaní normality

Rozsah výberu n	**20**	**50**	**100**	**200**	**500**
Teoretická hodnota ρ	-0,9000	-0,9000	-0,9000	-0,9000	-0,9000
r	-0,8943	-0,8952	-0,9004	-0,9026	-0,9012
r_s	-0,8548	-0,8715	-0,8848	-0,8919	-0,8915
Odhad r	-0,8656	-0,8813	-0,8938	-0,9004	-0,9001
Teoretická hodnota ρ	-0,1000	-0,1000	-0,1000	-0,1000	-0,1000
r	-0,0950	-0,0942	-0,0947	-0,1004	-0,0986
r_s	-0,0747	-0,0796	-0,0897	-0,0940	-0,0992
Odhad r	-0,0782	-0,0833	-0,0939	-0,0984	-0,1039
Teoretická hodnota ρ	0,2000	0,2000	0,2000	0,2000	0,2000
r	0,1838	0,1870	0,1982	0,1964	0,1997
r_s	0,1627	0,1792	0,1861	0,1835	0,1881
Odhad r	0,1702	0,1874	0,1946	0,1919	0,1966
Teoretická hodnota ρ	0,4000	0,4000	0,4000	0,4000	0,4000
r	0,3959	0,3857	0,4034	0,4090	0,3991
r_s	0,3529	0,3634	0,3835	0,3896	0,3827
Odhad r	0,3674	0,3783	0,3989	0,4052	0,3981
Teoretická hodnota ρ	0,8000	0,8000	0,8000	0,8000	0,8000
r	0,7870	0,7943	0,7951	0,8006	0,8017
r_s	0,7614	0,7681	0,7771	0,7801	0,7865
Odhad r	0,7764	0,7828	0,7915	0,7944	0,8006

Tabuľka A.4: Hodnoty koeficientov pri porušení normality jednou zložkou

Rozsah výberu n	**20**	**50**	**100**	**200**	**500**
Teoretická hodnota ρ	-0,5000	-0,5000	-0,5000	-0,5000	-0,5000
r	-0,4822	-0,4813	-0,4863	-0,4809	-0,4812
r_s	-0,4476	-0,4562	-0,4572	-0,4620	-0,4642
Odhad r	-0,4645	-0,4732	-0,4742	-0,4791	-0,4814
Teoretická hodnota ρ	-0,1000	-0,1000	-0,1000	-0,1000	-0,1000
r	-0,0661	-0,0811	-0,0908	-0,0970	-0,1070
r_s	-0,0748	-0,0864	-0,0951	-0,0968	-0,1022
Odhad r	-0,0783	-0,0904	-0,0995	-0,1013	-0,1070
Teoretická hodnota ρ	0,4000	0,4000	0,4000	0,4000	0,4000
r	0,3886	0,3936	0,3999	0,3967	0,3975
r_s	0,3601	0,3661	0,3820	0,3831	0,3793
Odhad r	0,3749	0,3810	0,3974	0,3985	0,3946
Teoretická hodnota ρ	0,8000	0,8000	0,8000	0,8000	0,8000
r	0,7899	0,7821	0,7806	0,7810	0,7770
r_s	0,7594	0,7626	0,7643	0,7653	0,7624
Odhad r	0,7744	0,7775	0,7792	0,7802	0,7774

Tabuľka A.5: Hodnoty koeficientov pri porušení normality oboma zložkami

Rozsah výberu n	**20**	**50**	**100**	**200**	**500**
Teoretická hodnota ρ	-0,9000	-0,9000	-0,9000	-0,9000	-0,9000
r	-0,8633	-0,8698	-0,8687	-0,8665	-0,8670
r_s	-0,8462	-0,8469	-0,8505	-0,8561	-0,8540
Odhad r	-0,8575	-0,8581	-0,8615	-0,8668	-0,8648
Teoretická hodnota ρ	-0,5000	-0,5000	-0,5000	-0,5000	-0,5000
r	-0,4624	-0,4794	-0,4817	-0,4766	-0,4833
r_s	-0,4430	-0,4562	-0,4590	-0,4615	-0,4683
Odhad r	-0,4598	-0,4732	-0,4761	-0,4786	-0,4855
Teoretická hodnota ρ	0,2000	0,2000	0,2000	0,2000	0,2000
r	0,1733	0,1844	0,1954	0,1968	0,1992
r_s	0,1643	0,1787	0,1821	0,1842	0,1890
Odhad r	0,1718	0,1868	0,1904	0,1926	0,1976
Teoretická hodnota ρ	0,8000	0,8000	0,8000	0,8000	0,8000
r	0,7831	0,7832	0,7811	0,7789	0,7790
r_s	0,7542	0,7664	0,7605	0,7602	0,7650
Odhad r	0,7694	0,7812	0,7755	0,7753	0,7799

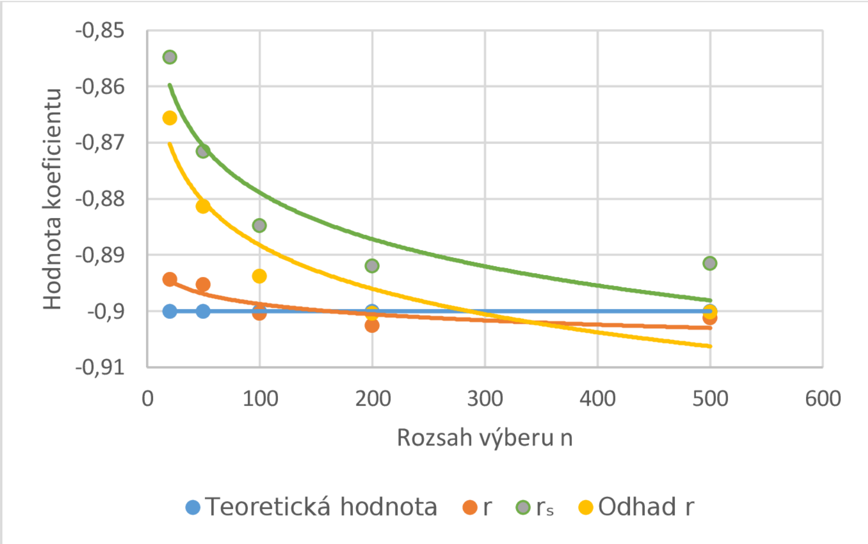
Tabuľka A.6: Hodnoty koeficientov pri výskyte odľahlých hodnôt

Rozsah výberu n	**20**	**60**	**100**	**200**	**500**
Teoretická hodnota ρ	-0,9000	-0,9000	-0,9000	-0,9000	-0,9000
r	-0,6621	-0,6642	-0,6715	-0,6873	-0,6813
r_s	-0,7887	-0,7978	-0,8013	-0,8091	-0,8111
Odhad r	-0,8026	-0,8113	-0,8147	-0,8221	-0,8241
Teoretická hodnota ρ	-0,5000	-0,5000	-0,5000	-0,5000	-0,5000
r	-0,3738	-0,3808	-0,3789	-0,3915	-0,3827
r_s	-0,4098	-0,4205	-0,4279	-0,4450	-0,4468
Odhad r	-0,4259	-0,4368	-0,4444	-0,4618	-0,4637
Teoretická hodnota ρ	0,2000	0,2000	0,2000	0,2000	0,2000
r	0,1285	0,1485	0,1429	0,1696	0,1587
r_s	0,1599	0,1801	0,1806	0,1893	0,1884
Odhad r	0,1673	0,1883	0,1889	0,1980	0,1970
Teoretická hodnota ρ	0,4000	0,4000	0,4000	0,4000	0,4000
r	0,2930	0,3067	0,3084	0,3045	0,3050
r_s	0,3262	0,3303	0,3544	0,3536	0,3530
Odhad r	0,3399	0,3442	0,3690	0,3682	0,3676

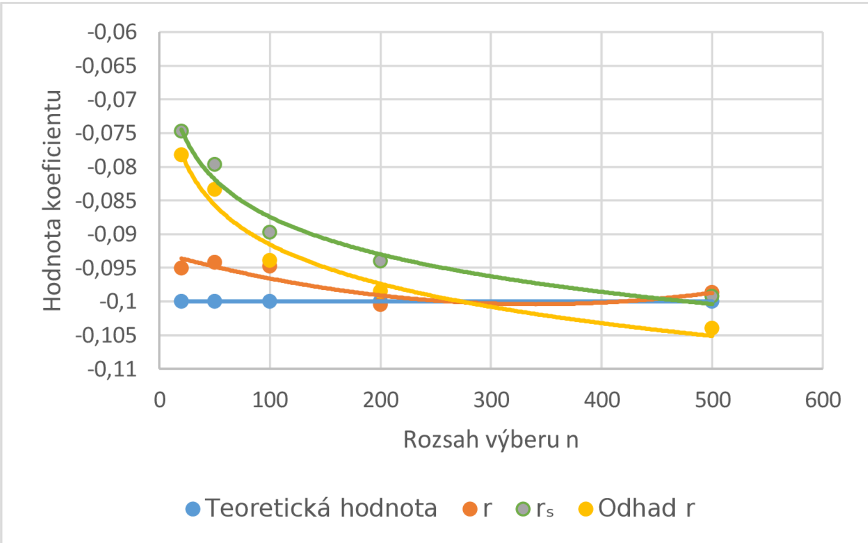
Tabuľka A.7: Hodnoty koeficientov pri skladaní rozdelení

Rozsah výberu n	**20**	**50**	**100**	**200**	**500**
$\rho_1 = 0{,}9;\ \rho_2 = -0{,}9;\ w_1 = 0{,}5;\ w_2 = 0{,}5$					
Teoretická hodnota ρ	0,0000	0,0000	0,0000	0,0000	0,0000
r	0,0997	0,0512	-0,0239	-0,0070	0,0045
r_s	0,0957	0,0621	-0,0252	-0,0044	0,0028
Odhad r	0,1002	0,0650	-0,0264	-0,0046	0,0030
$\rho_1 = 0{,}1;\ \rho_2 = -0{,}1;\ w_1 = 0{,}5;\ w_2 = 0{,}5$					
Teoretická hodnota ρ	0,0000	0,0000	0,0000	0,0000	0,0000
r	-0,0325	-0,0099	0,0014	0,0064	-0,0021
r_s	-0,0212	-0,0128	0,0019	0,0042	-0,0042
Odhad r	-0,0222	-0,0134	0,0020	0,0044	-0,0044
$\rho_1 = 0{,}45;\ \rho_2 = -0{,}8;\ w_1 = 0{,}4;\ w_2 = 0{,}6$					
Teoretická hodnota ρ	-0,3000	-0,3000	-0,3000	-0,3000	-0,3000
r	-0,4492	-0,4208	-0,4138	-0,4091	-0,4047
r_s	-0,4070	-0,3982	-0,3956	-0,3941	-0,3881
Odhad r	-0,4230	-0,4140	-0,4113	-0,4098	-0,4036

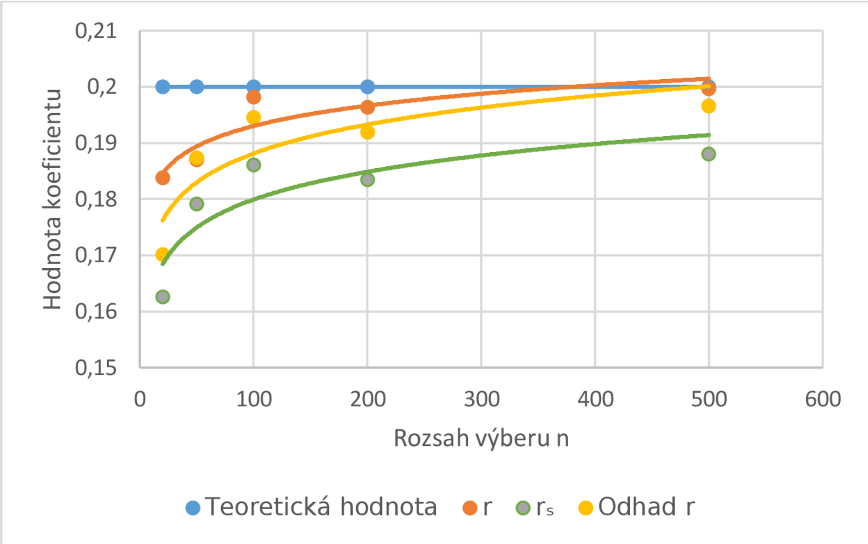
A.3 Grafy pre zvyšné simulácie



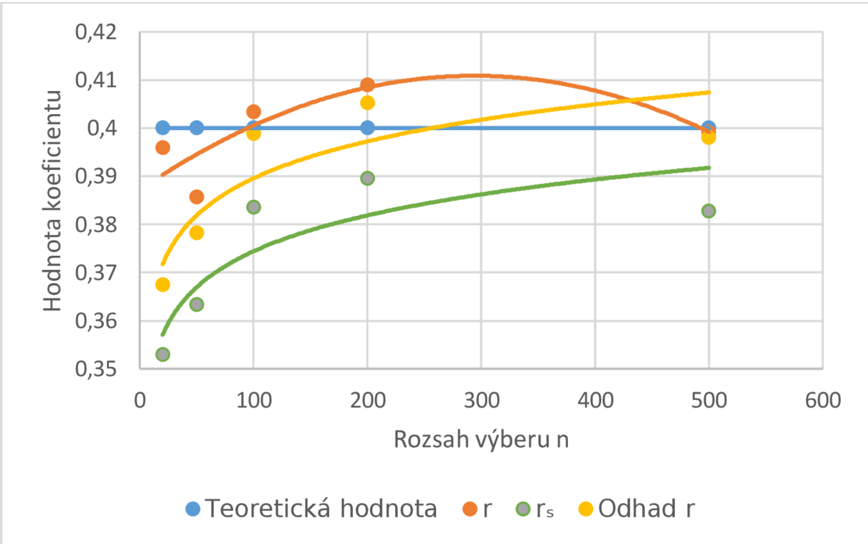
Obrázok A.8: Porovnanie koeficientov pre $\rho = -0{,}9$ pri zachovaní normality



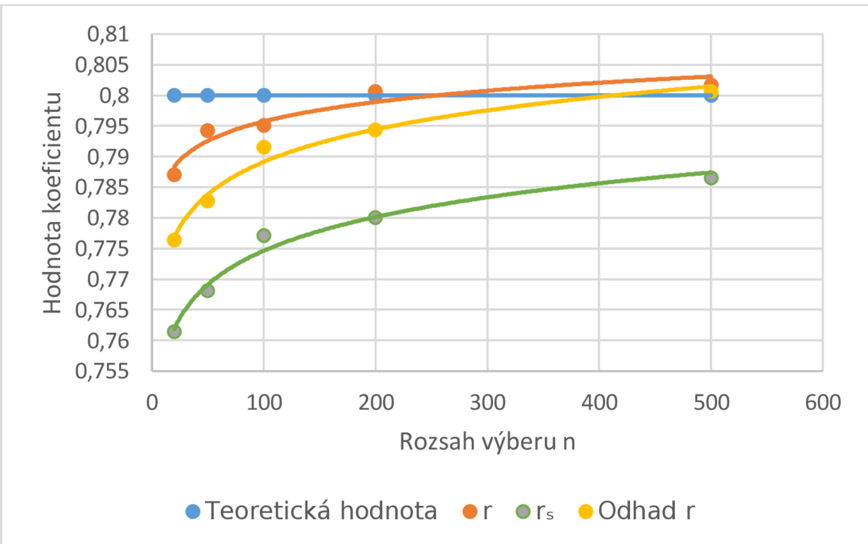
Obrázok A.9: Porovnanie koeficientov pre $\rho = -0{,}1$ pri zachovaní normality



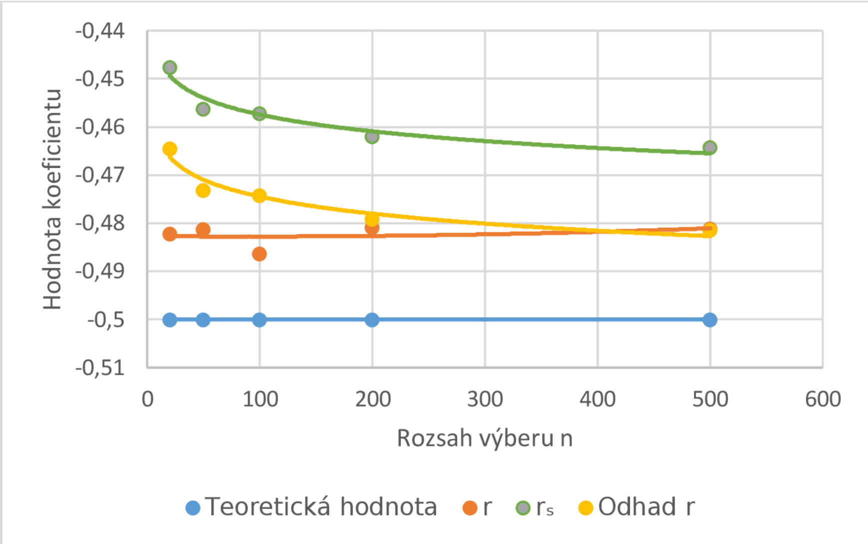
Obrázok A.10: Porovnanie koeficientov pre $\rho = 0{,}2$ *pri zachovaní normality*



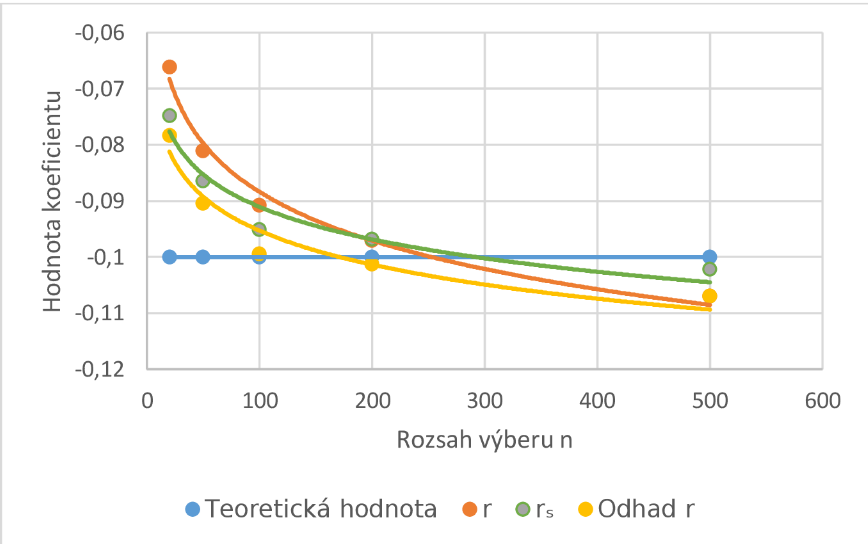
Obrázok A.11: Porovnanie koeficientov pre $\rho = 0{,}4$ *pri zachovaní normality*



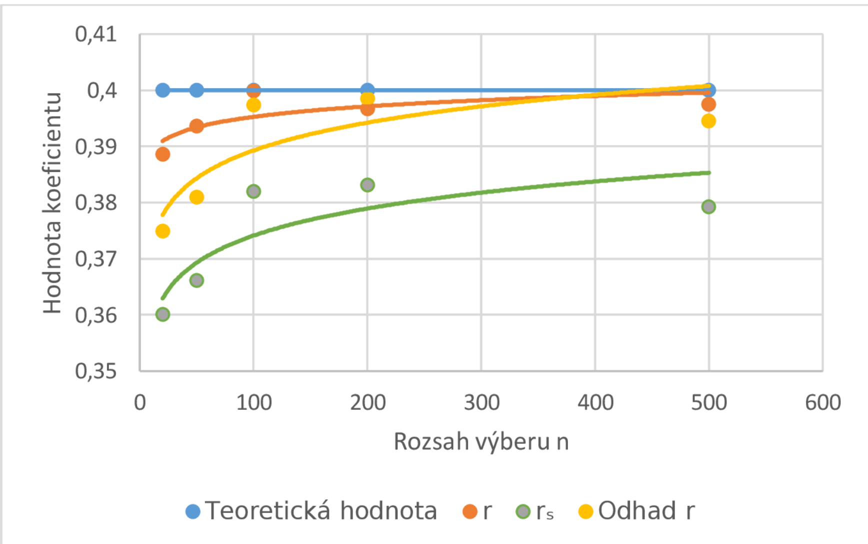
Obrázok A.12: Porovnanie koeficientov pre $\rho = 0{,}8$ *pri zachovaní normality*



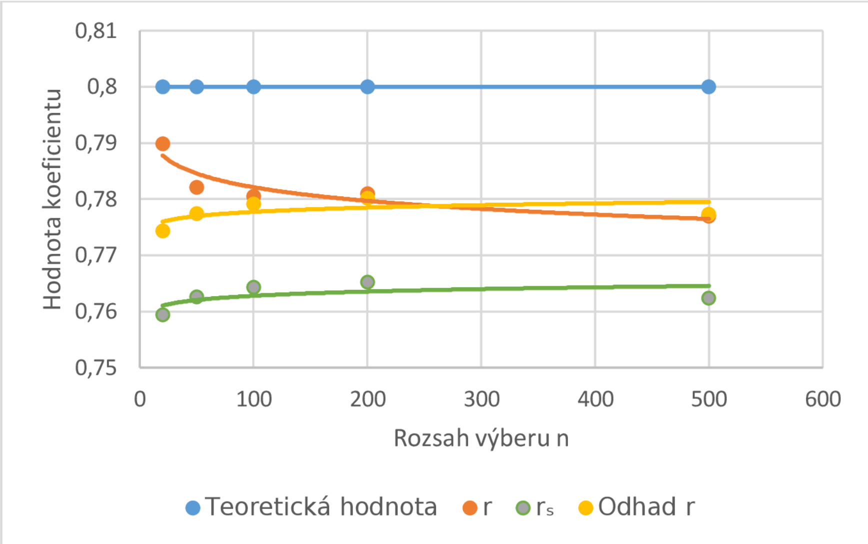
Obrázok A.13: Porovnanie koeficientov pre $\rho = -0{,}5$ *pri porušení normality jednou zložkou*



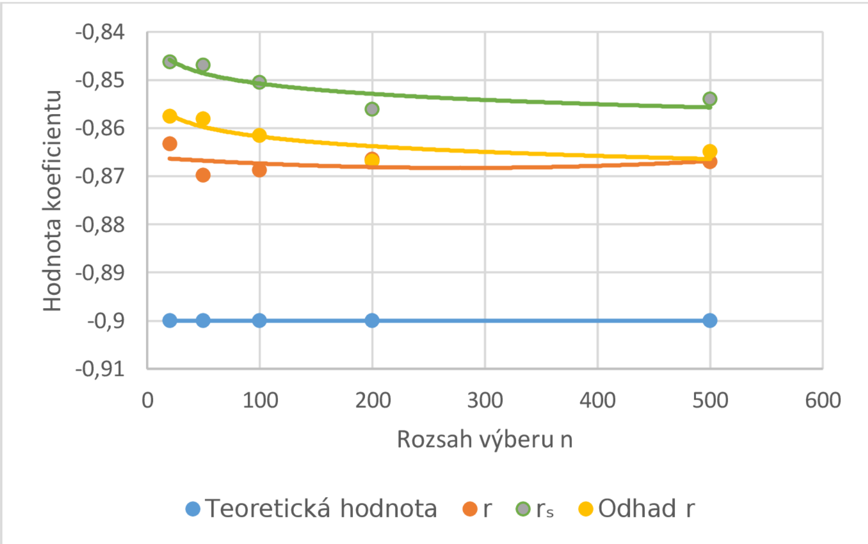
Obrázok A.14: Porovnanie koeficientov pre $\rho = -0{,}1$ *pri porušení normality jednou zložkou*



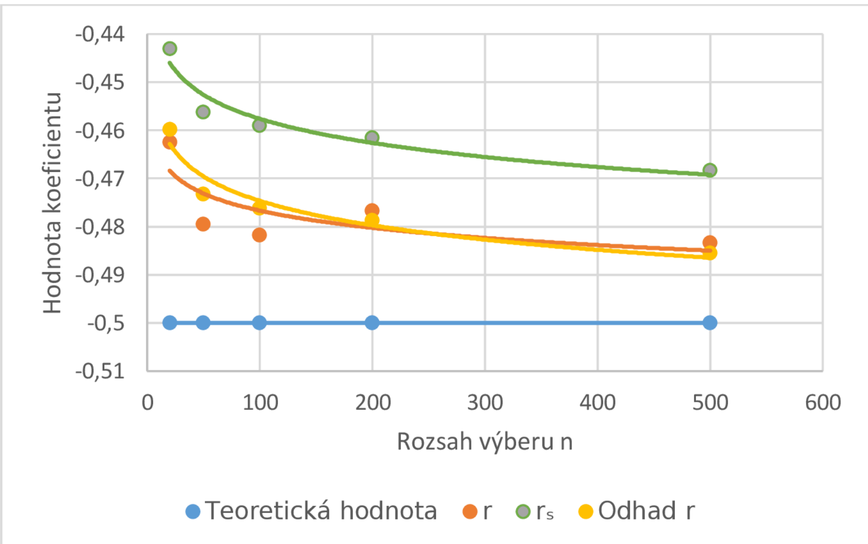
Obrázok A.15: Porovnanie koeficientov pre $\rho = 0{,}4$ *pri porušení normality jednou zložkou*



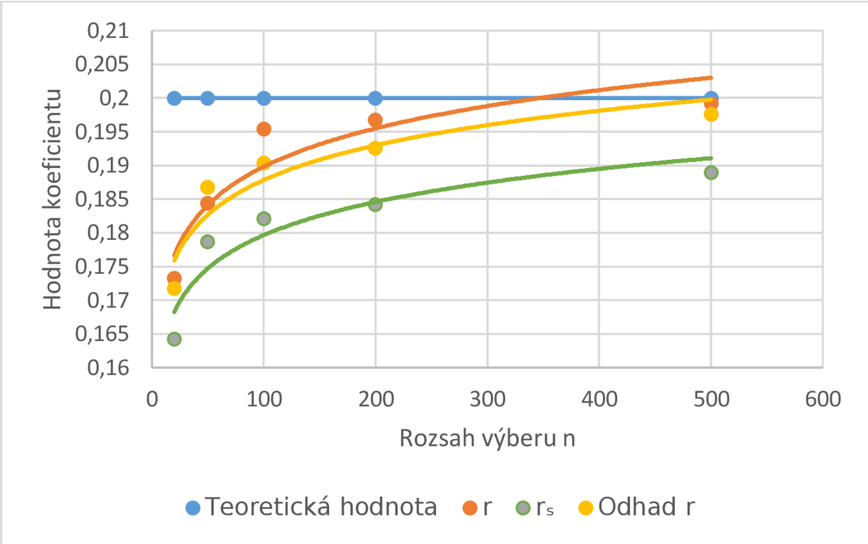
Obrázok A.16: *Porovnanie koeficientov pre* $\rho = 0{,}8$ *pri porušení normality jednou zložkou*



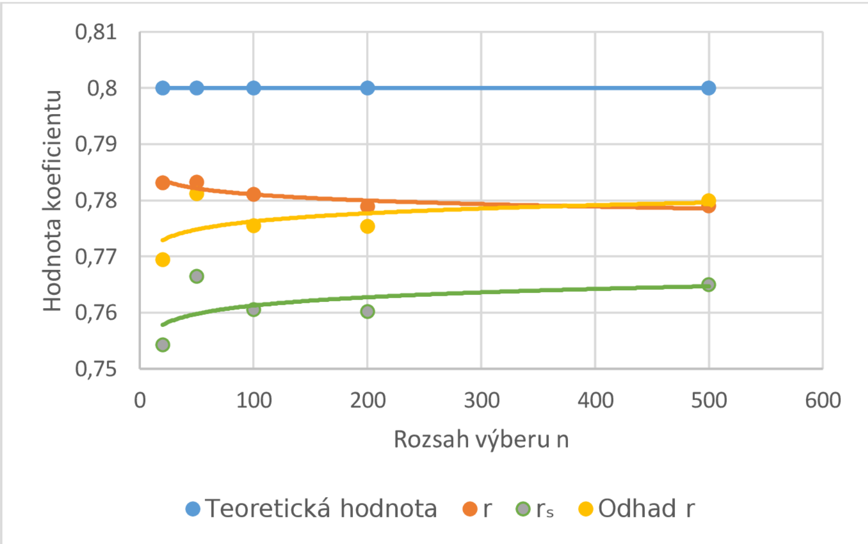
Obrázok A.17: *Porovnanie koeficientov pre* $\rho = -0{,}9$ *pri porušení normality oboma zložkami*



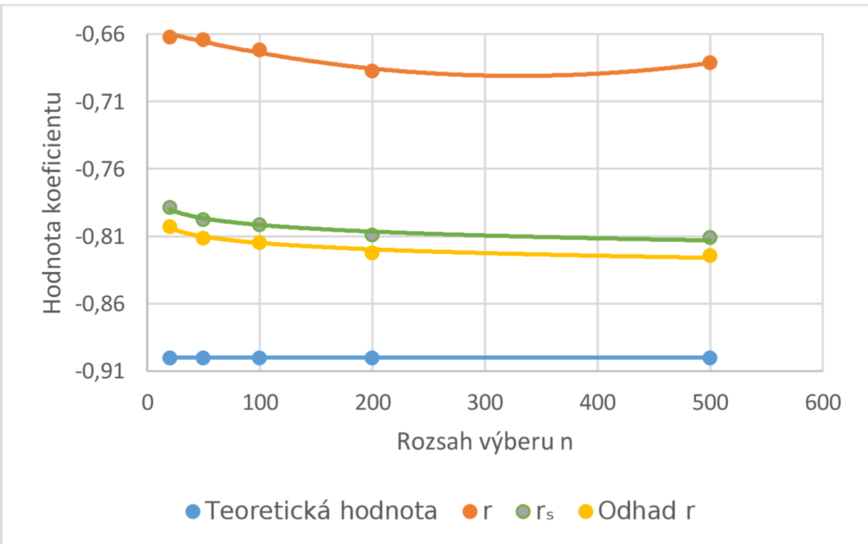
Obrázok A.18: *Porovnanie koeficientov pre* $\rho = -0{,}5$ *pri porušení normality oboma zložkami*



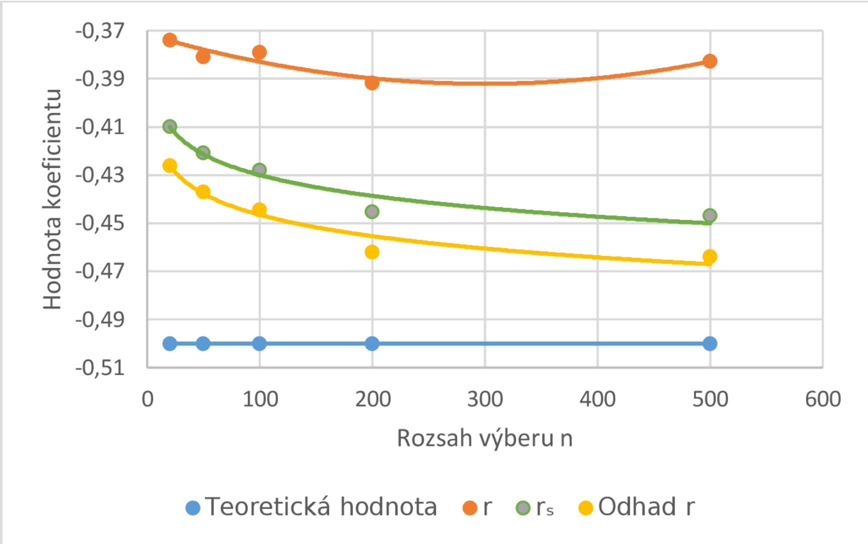
Obrázok A.19: Porovnanie koeficientov pre $\rho = 0{,}2$ *pri porušení normality oboma zložkami*



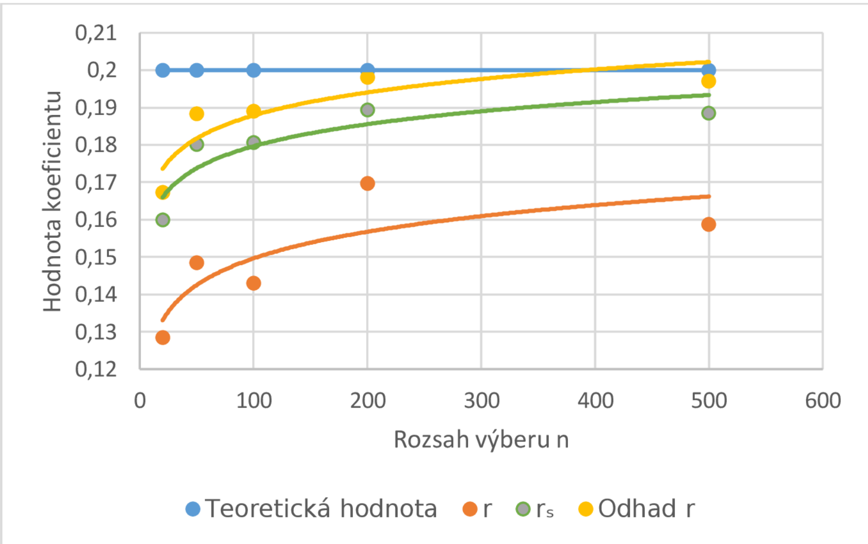
Obrázok A.20: Porovnanie koeficientov pre $\rho = 0{,}8$ *pri porušení normality oboma zložkami*



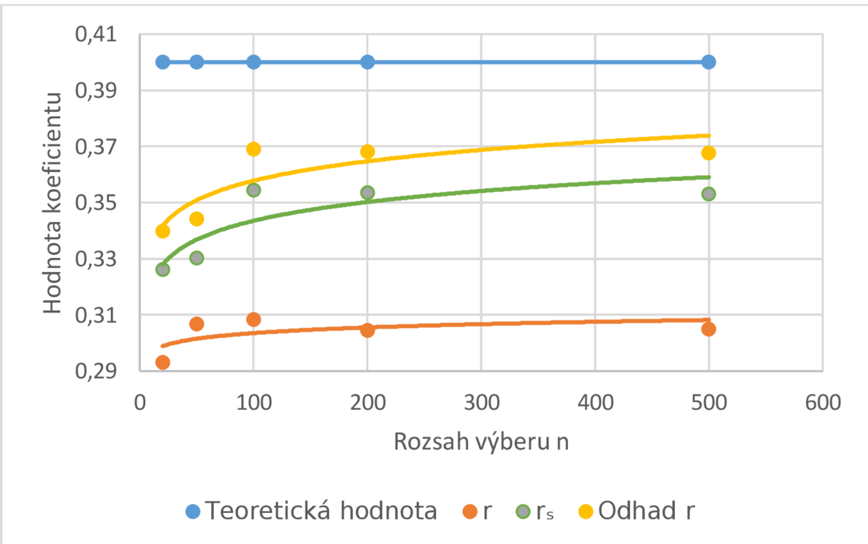
Obrázok A.21: Porovnanie koeficientov pre $\rho = -0{,}9$ *pri výskyte odľahlých hodnôt*



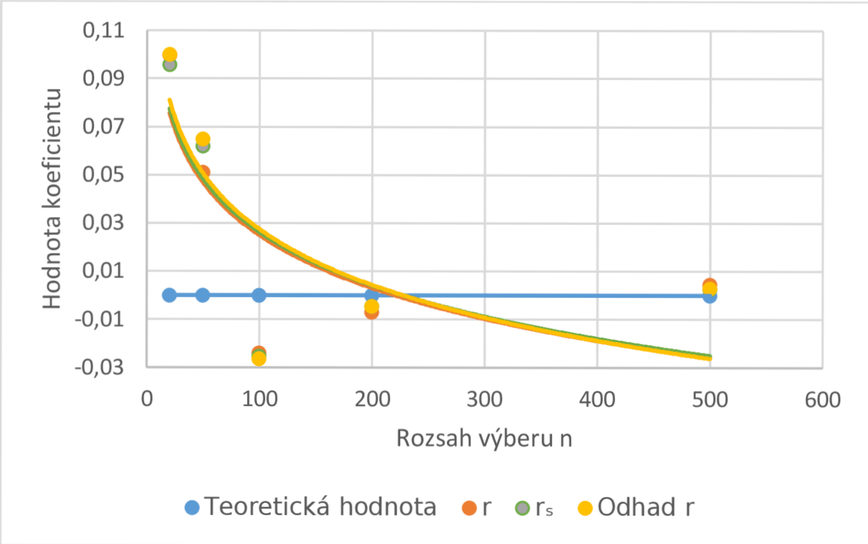
Obrázok A.22: Porovnanie koeficientov pre $\rho = -0{,}5$ *pri výskyte odľahlých hodnôt*



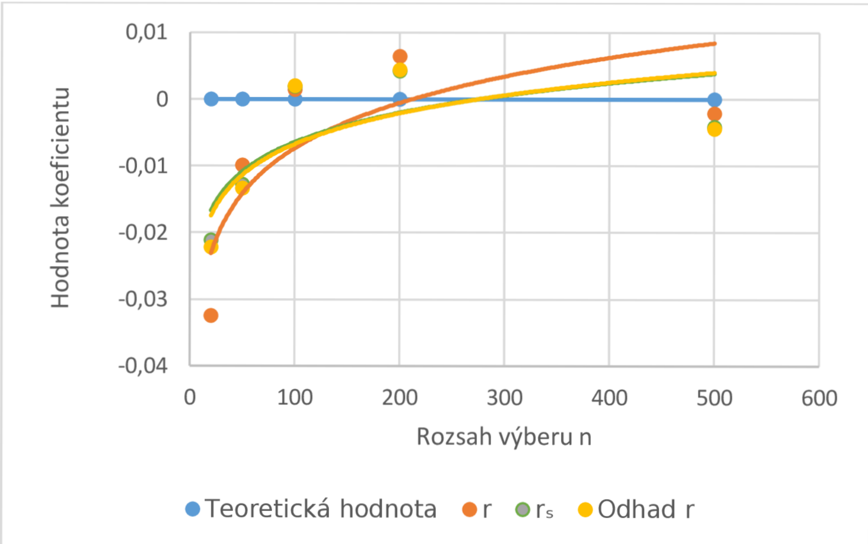
Obrázok A.23: Porovnanie koeficientov pre $\rho = 0{,}2$ *pri výskyte odľahlých hodnôt*



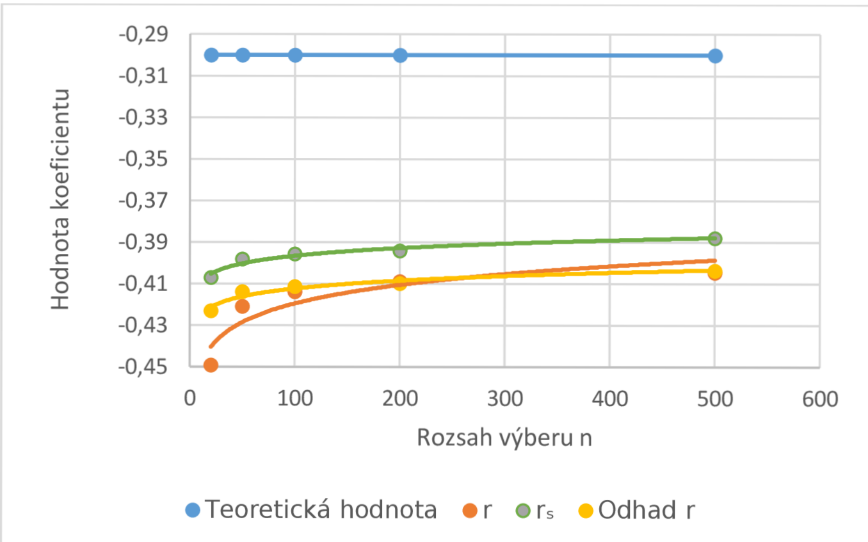
Obrázok A.24: Porovnanie koeficientov pre $\rho = 0{,}4$ *pri výskyte odľahlých hodnôt*



Obrázok A.25: Porovnanie koeficientov pre $\rho_1 = 0{,}9;\ \rho_2 = -0{,}9;\ w_1 = 0{,}5;\ w_2 = 0{,}5$



Obrázok A.26: Porovnanie koeficientov pre $\rho_1 = 0{,}1;\ \rho_2 = -0{,}1;\ w_1 = 0{,}5;\ w_2 = 0{,}5$



Obrázok A.27: Porovnanie koeficientov pre $\rho_1 = 0{,}45;\ \rho_2 = -0{,}8;\ w_1 = 0{,}4;\ w_2 = 0{,}6$

B Obsah priloženého CD

Súčasťou priloženého CD sú:

Hodnoty_korelácia	Súbor so získanými hodnotami korelačných koeficientov
dvojroz_norm	Zdrojový kód k podkapitole 3.1
dvojroz_norm_weib	Zdrojový kód k podkapitole 3.2
dvojroz_norm_posun_y	Zdrojový kód k podkapitole 3.3
dvojroz_norm_cez_vahy	Zdrojový kód k podkapitole 3.4
Bakalarska_praca	PDF dokument s bakalárskou prácou.