Kappa, jako všechny koeficienty pro pořadové znaky, nabývá rozsahu hodnot od −1 do +1. Hodnota 1 znamená perfektní souhlas, hodnota 0 znamená takový počet shod, který odpovídá náhodné shodě pozorovatelů (Hendl, 2004, s. 232), hodnota −1 perfektní nesouhlas. Hodnotu koeficientu *kappa* pro naši situaci ukazuje výstup 9.5b a očividně její hodnota 0,17 značí velmi nízkou míru souhlasu – de facto se jedná o hodnotu, která se blíží náhodnému rozložení na diagonále. Česká veřejnost se tedy při vyjadřování důvěry poslancům a senátorům ve svých postojích neshodu-je. Z frekvencí jednotlivých proměnných nebo jejich průměrů pak vyplývá, že o něco vyšší důvěru požívala Poslanecká sněmovna.

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Measure of Agreement	Kappa	,165	,018	10,064	,000
N of Valid Cases		1748			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Výstup 9.5b Koeficient souhlasu pro důvěru v Parlament a Senát

V literatuře se vyskytl návod, jak interpretovat hodnoty tohoto koeficientu: *kappa*, která je nižší než 0,40, indikuje velmi nízkou míru souhlasu (Fleiss, 1981, s. 218). Je to však návod, který nemůže být brán příliš rigidně. Jedním z důvodů je, jak nabádají Sim a Wright (2005), že čím vyšší je počet kategorií sledované proměnné, tím vyšší je možnost, že dojde k nesouhlasu v kategoriích, což se projeví tak, že *kappa* bude mít nižší hodnotu u proměnné s více kategoriemi než u proměnné s méně kategoriemi. Ve výstupu 9.5b si všimněme, že SPSS tiskne i signifikanci tohoto koeficientu, statistikové však upozorňují, že testovat nulovou hypotézu nemá v případě úloh o shodě smysl.

9.5 Míra souvislosti pro intervalové znaky

Souvislost mezi dvěma znaky intervalovými se měří prostřednictvím jednoho jediného koeficientu – **Pearsonova koeficientu lineární korelace** (značíme symbolem ***r***). Vypočítá se prostřednictvím kovariance, to je variance (rozptylu) pro dvě proměnné, kdy každou odchylku od průměru jedné proměnné ($x_i - \bar{x}$) násobíme odchylkou od průměru druhé proměnné ($y_i - \bar{y}$), tyto odchylky sečteme a podělíme $N - 1$:

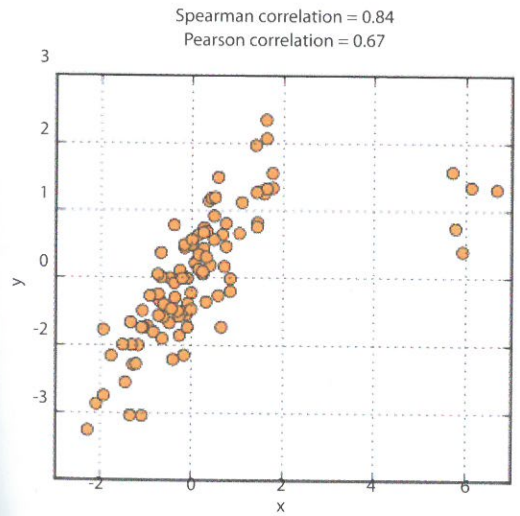
$$\mathrm{cov}(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N-1}.$$

Tuto kovarianci standardizujeme (podělíme součinem směrodatných odchylek obou proměnných). Vzorec má pak tuto podobu:

$$r = \frac{\mathrm{cov}_{xy}}{s_x\, s_y}.$$

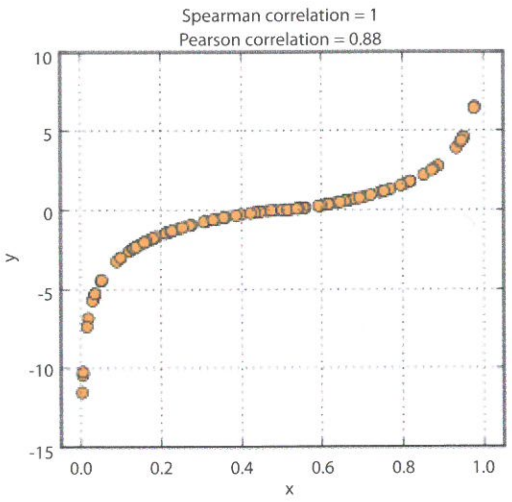
Pearsonův koeficient je koeficientem symetrickým, nabývá hodnot od −1 do +1 a fakt, že je to koeficient lineární korelace znamená, že dokáže zachytit a změřit pouze lineární vztah mezi dvěma proměnnými. Lineární vztah nastává tehdy, když s měnícími se hodnotami jedné proměnné se proporcionálně mění hodnoty druhé proměnné – kdybychom hodnoty obou proměnných zaznamenali do dvourozměrného bodového grafu, body budou ležet na přímce. Pokud zjistíme, že Pearsonův koeficient má hodnotu 0, neznamená to ještě, že mezi sledovanými proměnnými není vztah – nula zde indikuje, že vztah nemá lineární podobu (totéž indikuje nula u Spearmanova koeficientu). Jelikož je výpočet korelačního koeficientu r založen na rozptylu (a tudíž odchylkách od průměru), je jeho hodnota citlivá na odlehlé hodnoty – pro srovnání Spearmanův koeficient, který se vypočítává podle stejné rovnice jako r, kdy však hodnoty proměnných x a y tvoří pořadí, na odlehlé hodnoty citlivý není.

Ilustruje to následující obrázek (viz obr. 9.4), na němž je pro dané rozložení srovnávána hodnota korelace Pearsonova a Spearmanova: Pearsonova korelace ($r = 0{,}67$) je nižší než korelace Spearmanova ($r_s = 0{,}84$). Rozdíl je způsoben odlehlými hodnotami v pravé části obrázku.



Obr. 9.4 Pearsonova a Spearmanova korelace pro data s odlehlými hodnotami

A ještě jedno zajímavé srovnání si uveďme: na obr. 9.5 je podobné srovnání hodnot těchto dvou koeficientů v případě, kdy vztah má esovitou, to je nelineární, avšak monotónní podobu. Spearmanův koeficent má hodnotu $r_s = 1{,}00$, zatímco Pearsonovo $r = 0{,}88$. Tyto příklady mimochodem pěkně ukazují, jak je důležité si při měření souvislosti u pořadových nebo intervalových proměnných nechat udělat bodový graf, z něhož zjistíme, jaký má hledaný vztah průběh – což je důležité jak pro nasazení příslušného koeficientu, tak také pro jeho interpretaci.

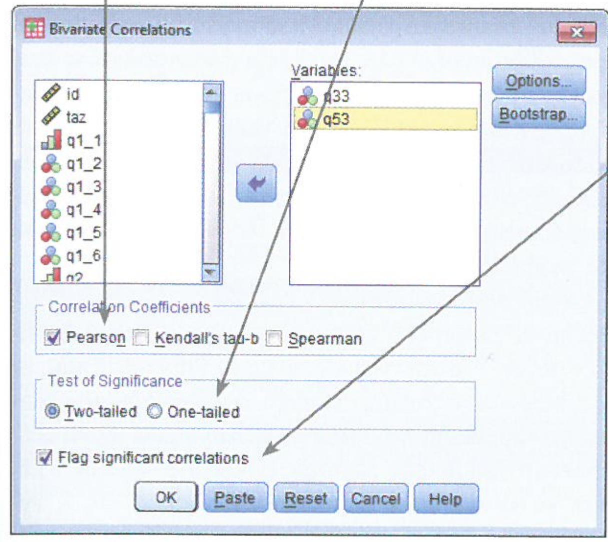


Obr. 9.5 Pearsonova a Spearmanova korelace pro data s odlehlými hodnotami

Pearsonův koeficient korelace zpravidla vypočítáváme v SPSS pomocí jiné procedury než předchozí koeficienty.[222] Z předcházejících kapitol již víme, že intervalové znaky se mimo jiné vyznačují tím, že mají dlouhé stupnice měření (např. proměnná věk má u dospělých respondentů více než 60 kategorií, příjem může mít desetitisíce kategorií, levo-pravá politická orientace může mít deset kategorií atd.). Bylo by proto nesmyslné nechat vytvářet pro takovéto znaky tabulku třídění II. stupně (*Crosstabs*). Například kdybychom třídili proměnnou *levo-pravá politická orientace* měřenou na desetibodové stupnici s proměnnou *důležitost Boha* v životě jedince měřenou rovněž na desetibodové stupnici, vznikne tabulka o 10 sloupcích a 10 řádcích (takže by měla 100 políček), která se nedá jednak smysluplně zobrazit ve výstupu, jednak se nedá ani smysluplně interpretovat. Z tohoto důvodu má SPSS nastavenou možnost vypočítat Pearsona (ale i Spearmana a Kendallovo *tau*) bez tabulky *Crosstabs*. Je jí procedura *Correlate*, která tiskne jako výstup **matici korelací**.

222 Jistě jste si v obr. 9.1b všimli, že Pearsonův koeficient je zabudován i v proceduře *Crosstabs*. Tento způsob výpočtu má smysl použít tehdy, když intervalové proměnné mají krátké stupnice měření. Což by např. bylo v případě, kdybychom hledali souvislost mezi počtem dětí (hodnoty této proměnné se obvykle pohybují od 0 do 4) a mírou anomie (tato stupnice nabývá hodnot od 0 do 5). Ale ani při *Crosstabsu* není třeba tabulky vytvářet. Když zaškrtneme požadavek *Supress tables*, objeví se pouze požadované koeficienty.

Procedura: *Analyze – Correlate – Bivariate* – proměnné, jejichž vztahy hledáme – volba koeficientů – volba jedno či dvoustranného testu signifikance – zvýrazni signifikantní korelace.



Obr. 9.6 Dialogové okno pro zadání výpočtu matice korelací

Příklad 9.4

Existuje v datech EVS statistická souvislost mezi politickou orientací měřenou na levo-pravém kontinuu (*q53*) a názorem na důležitost Boha v životě jedince (*q33*)?

Řešení: Zadání tohoto výpočtu ukazuje obr. 9.6. Výstup vypadá následovně (viz výstup 9.6). Výsledná matice korelací má vždy podobu čtvercové tabulky obsahující tolik řádků a sloupců, kolik proměnných vstupuje do analýzy. Všimněte si, že korelace proměnných se sebou samými jsou umístěny na diagonále tabulky a jsou vždy rovny 1.

Correlations

		Q33 Bůh - důležitost v životě	Q53 Levice - pravice
Q33 Bůh - důležitost v životě	Pearson Correlation	1,000	,147**
	Sig. (2-tailed)	.	,000
	N	1846	1711
Q53 Levice - pravice	Pearson Correlation	,147**	1,000
	Sig. (2-tailed)	,000	.
	N	1711	1758

**. Correlation is significant at the 0.01 level (2-tailed).

Pozn. Žluté zvýraznění políčka je provedeno z didaktických důvodů, SPSS to sám o sobě nedělá.

Výstup 9.6