

Odpovědi (téma 5)

1. c

2. a

3. d

4. cdeab

5.1 omezená variabilita snižuje výsledný koeficient a může vést k podhodnocení korelace v populaci

5.2 rozdílná rozložení proměnných snižují výsledný koeficient

5.3 odlehle hodnoty se velmi silně promítají do výsledku a mohou neadekvátně zvýšit (v případě, že leží poblíž regresní přímky) či snížit (pokud leží kdekoli jinde) výsledný koeficient

5.4 použití extrémních skupin může zvýšit hodnotu korelace

6. a) pozitivní, b) nulová, c) pozitivní, ~~d) nulová~~, e) pozitivní (více hodin - asi vyšší číselná hodnota známky), f) nulová, g) nulová, h) negativní, i) pozitivní

7. 0,8

8. W

9. nijak

10. v prvním případě $z_Y = -0,5$; ve druhém $z_Y = -0,8$

11. c – v případě prvních dvou korelací je rozdíl mezi vysvětleným rozptylem $0,4^2 - 0,2^2 = 0,16 - 0,04 = 0,12$, tj. 12 %; u druhé dvojice je rozdíl $0,7^2 - 0,5^2 = 0,49 - 0,25 = 0,24$, tj. 24 %

12. 1,0

13. Lineárními transformacemi (přičítání, násobení konstantou) nejsou ovlivněna relativní pořadí skóre ani relativní rozdíly mezi skóre, a tedy ani korelace. Velikost korelace zde tedy na jednotkách nezáleží. Oběma badatelům vyšla nakonec stejná korelace.

14. Maximální korelace může být 1, tedy $\max(c_{XY}) = s_X s_Y$, tedy 45.

15. Korelace na naměřených datech bude vyšší, protože data budou zatížena menší chybou měření.

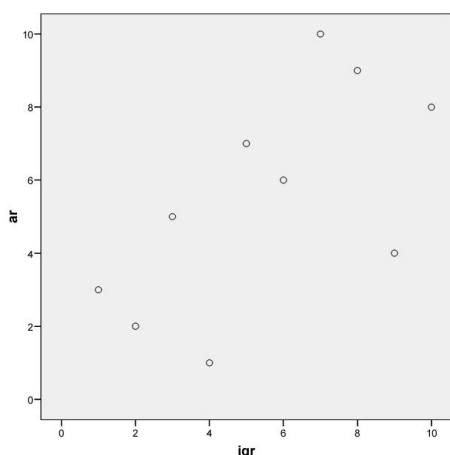
16. c

17. Korelace by byla podstatně větší.

18. Reliabilita kritéria (tedy přesnost hodnocení) je mizerná. Zdá se, že každý hodnotitel hodnotí něco jiného. Korelace mezi výsledky a schopností učit tedy bude ve skutečnosti zřejmě vyšší.

19. a) $r = 0,52$

b) viz scatter, ano, ano



c) $m_{iq}=112$ a to je o dost víc než v populaci, $s_{iq}=15,7$ a to je jen nepatrně víc než populaci. Korelace by tím neměla být ovlivněna.

20. Odhad výšky korelace, směr korelace (pozitivní/negativní apod.), lineárnost a přítomnost odlehlých hodnot.

21. $r = 0.1288 / 3 = 0,043$

22.

5. a. $r = 0$
- b. $r = +.50$
- c. $r = -.50$
- d. $r = +.90$

23.

6.	r_{crit} 5%	r_{crit} 1%	r_{crit} 5%	r_{crit} 1%
a.	.388	.496	reject	reject
b.	.388	.496	reject	accept
c.	.532	.661	reject	reject
d.	.444	.561	reject	reject
e.	.355	.456	reject	accept
f.	.666	.798	reject	reject
g.	.514	.641	reject	accept
h.	.195	.254	reject	accept

24.

7. a. $\Sigma X = 24$ $\Sigma Y = 21$ $\Sigma XY = 125$
 $(\Sigma X)^2 = 576$ $(\Sigma Y)^2 = 441$ $n_p = 4$
 $\Sigma X^2 = 150$ $\Sigma Y^2 = 113$

$$r_{obt} = \frac{-4}{16.25} = -.25$$

- b. $H_0: \rho = 0; H_1: \rho \neq 0$
- c. $r_{crit} (2) = .950$
- d. Do not reject the null hypothesis.
- e. $r^2 = (-.25)^2 = .0625$ or 6.25% (However, r^2 would not be reported since r is nonsignificant.)

25.

8. a. $\Sigma X = 32$ $\Sigma Y = 25$ $\Sigma XY = 213$
 $(\Sigma X)^2 = 1024$ $(\Sigma Y)^2 = 625$ $n_p = 4$
 $\Sigma X^2 = 266$ $\Sigma Y^2 = 175$

$$r_{obt} = \frac{52}{54.78} = .95$$

- b. $H_0: \rho = 0; H_1: \rho \neq 0$
- c. $r_{crit} (2) = .950$
- d. Reject the null hypothesis when r is equal to or greater than r_{crit} .
- e. $r^2 = (.95)^2 = .90$, or 90%

26.

9. a. $\Sigma X = 43$ $\Sigma Y = 41$ $\Sigma XY = 453$
 $(\Sigma X)^2 = 1849$ $(\Sigma Y)^2 = 1681$ $n_p = 4$
 $\Sigma X^2 = 471$ $\Sigma Y^2 = 449$

$$r_{\text{obt}} = \frac{49}{63.44} = .77$$

- b. $H_0: \rho = 0; H_1: \rho \neq 0$
c. $r_{\text{crit}}(2) = .950$
d. Do not reject the null hypothesis.
e. 59.3 percent (r^2 would not be reported since r is nonsignificant.)

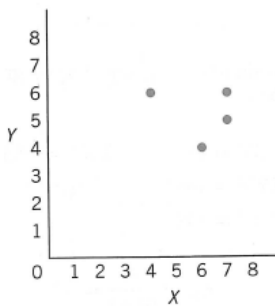
27.

12. a. $\Sigma X = 18$ $\Sigma Y = 30$ $\Sigma XY = 96$
 $(\Sigma X)^2 = 324$ $(\Sigma Y)^2 = 900$ $n_p = 5$
 $\Sigma X^2 = 80$ $\Sigma Y^2 = 190$

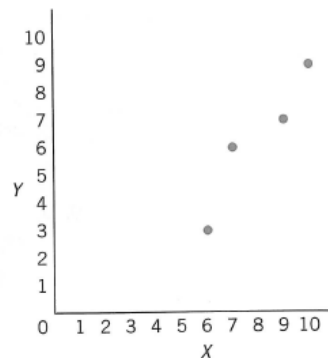
$$r_{\text{obt}} = \frac{-60}{61.64} = -.97$$

- b. $H_0: \rho = 0; H_1: \rho \neq 0$
c. $r_{\text{crit}}(3) = .878$
d. Reject the null hypothesis.
e. 94.10 percent.

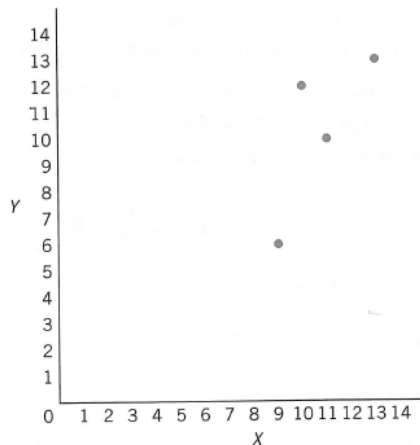
28. Grafy:



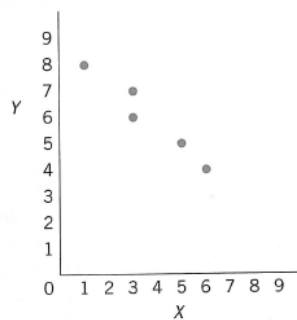
(Problem 7)



(Problem 8)



(Problem 9)



(Problem 12)

29. $r = 64 / 74 = 0,86$

30. r bude podhodnocovat ρ .

31. a 32.

18. a. $\Sigma X = 56$ $\Sigma Y = 503$ $\Sigma XY = 3887$
 $(\Sigma X)^2 = 3136$ $(\Sigma Y)^2 = 253,009$ $n_p = 7$
 $\Sigma X^2 = 542$ $\Sigma Y^2 = 36,365$

$$r_{\text{obt}} = \frac{-959}{1008.60} = -.95$$

b. $H_0: \rho = 0; H_1: \rho \neq 0$

c. $r_{\text{crit}}(5) = .754$

d. 90.25 percent

e. Reject the null hypothesis.

19. a. $\Sigma X = 18$ $\Sigma Y = 621$ $\Sigma XY = 2022$
 $(\Sigma X)^2 = 324$ $(\Sigma Y)^2 = 385,641$ $n_p = 6$
 $\Sigma X^2 = 94$ $\Sigma Y^2 = 64,979$

$$r_{\text{obt}} = \frac{954}{1007.93} = .95$$

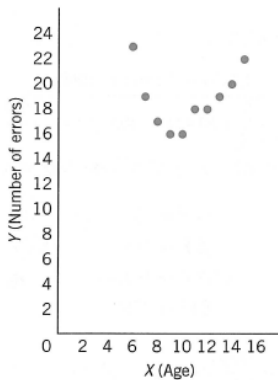
b. $H_0: \rho = 0; H_1: \rho \neq 0$

c. $r_{\text{crit}}(4) = .811$

d. 90.25 percent

e. Reject the null hypothesis.

33. Graf:



Pearsonův korelační koeficient není vhodný, neboť zohledňuje pouze lineární závislost.

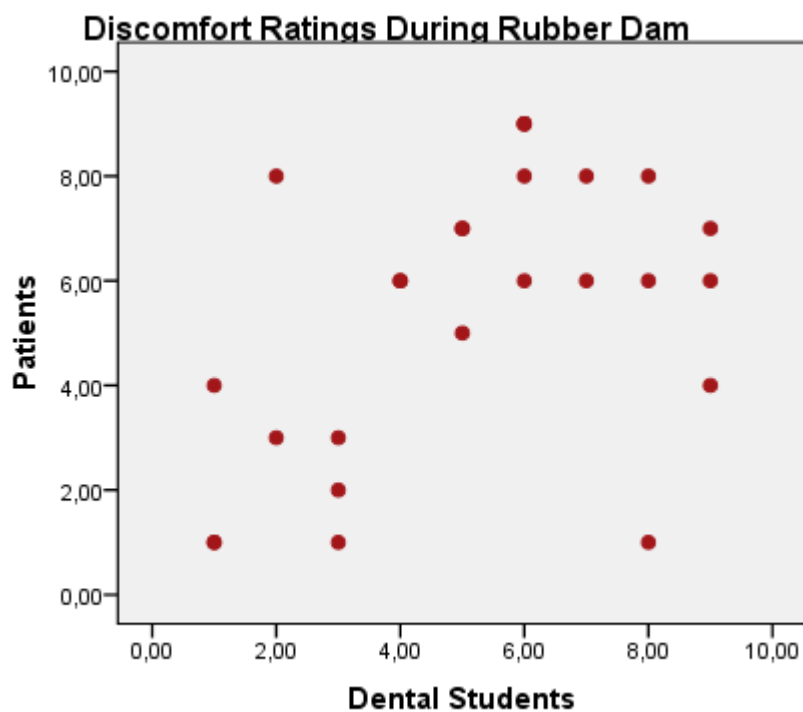
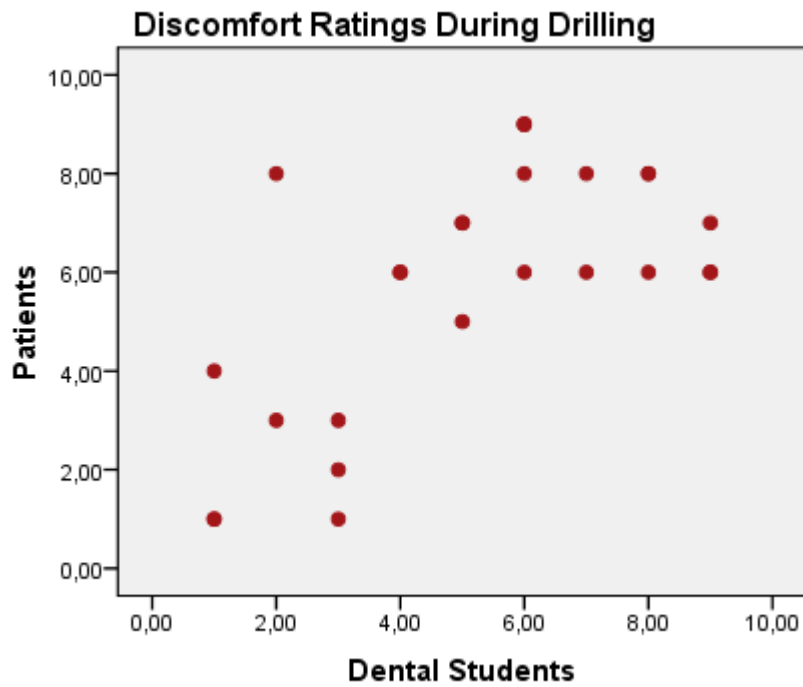
34. a 35.

21. Drilling: $r(24) = .65, p < .05$

Rubber Dam: $r(24) = .46, p < .05$

22. $r(21) = .413, p < .05$

Scattery ke 34:



36. a) $r = 0,011$ (kdo si to zkontroloval, zjistí, že průměr sebehodnocení není 0) ; b) mezi proměnným neexistuje žádný vztah; c) 0,01%

37. zvyšuje: 2 outlieři vpravo nahoře; snižuje: efekt podlahy u obou proměnných

38. τ (tau)

Konkordance: / / / / : 4

Diskordance: /// // / : 6

$\tau = (K-D)/\text{počet srovnání} = -0,2$

39. celkové četnosti hodnot jednotlivé proměnné, v kontingenční tabulce je uvádíme v posledním sloupci a posledním řádku

40. vzorek s příliš nízkou variabilitou

41. d – horní graf; a – dolní graf; Spearmanův koeficient by byl pravděpodobně vyšší, protože není tak necitlivý k nelineárnímu vztahu proměnných

42. $\tau = 0,6$

43.1 Hodnoty shrnuje tabulka:

	průměr	rozptyl	směr. odchylka
úzkost ze smrti	20,27	133,64	11,56
religiozita	10,93	26,21	5,12

Abychom však mohli tyto hodnoty vypočítat, musíme předpokládat alespoň intervalovou úroveň měření.

43.2 Kovariance je -41,2 a pearsonův korelační koeficient -0,696.

43.3 Korelační koeficient -0,696 naznačuje poměrně těsný negativní vztah, tedy s rostoucí religiozitou klesá úzkost ze smrti. Ovšem na základě korelace není nijak možné usuzovat na kauzalitu tohoto vztahu.

43.4 c

44.1 Hodnoty shrnuje tabulka:

	průměr	směr. odchylka
x	1,00	4,47
y	19,67	19,36

44.2 Kovariance je 40,0 a pearsonův korelační koeficient 0,462. I přesto, že mezi proměnnými existuje zřetelný a přesný vztah, korelační koeficient poukazuje na mnohem menší vztah. To je omezení pearsonova korelačního koeficientu, je použitelný pouze pro lineární vztahy.

45.1 Scatter A. Pearsonův korelační koeficient je -0,996, jde tedy o velice těsný negativní vztah.

45.2 Scatter D. Korelační koeficient 0,696, jedná se o poměrně silný pozitivní vztah.

45.3 Scatter B a C. Korelační koeficient je -0,002 (tedy prakticky 0) u B, -0,152 u C.

46.1 Hodnoty shrnuje tabulka:

	průměr	rozptyl	směr. odchylka
výška v palcích	67	42,5	6,52
self-efficacy	4	0,3	0,55

Kovariance je 3,25 a pearsonův korelační koeficient 0,91. Mezi výškou v palcích a self-efficacy podle tohoto výzkumu existuje velmi těsný pozitivní vztah.

46.2 Hodnoty druhého výzkumníka jsou také shrnuty v tabulce:

	průměr	rozptyl	směr. odchylka
výška v cm	167,5	265,6	16,3
sebehodnocení – druhý test	20	7,5	2,74

Kovariance je 40,625 a pearsonův korelační koeficient 0,91.

46.3 Průměr a směrodatná odchylka výšky se zvětšily 2,5x (tedy stejně jako data), rozptyl výšky 6,25x (= $2,5^2$). Průměr a směrodatná odchylka self-efficacy se zvýšily 5x, rozptyl 25x (= 5^2). Kovariance se znásobila součinem obou koeficientu, tedy (2,5 x 5)-krát. Korelace se nezměnila. Lineární transformace dat nemají na pearsonův korelační koeficient vliv. Například tedy nezáleží, v jakých jednotkách data měříme.

47.1 Kontingenční tabulka:

	fyzika			
matematika	1	2	3	Celkový

				součet
1	3	3	1	7
2	3	3	0	6
3	0	2	3	5
4	0	0	2	2
Celkový součet	6	8	6	20

Lineární vztah se projevuje vysokými četnostmi na jedné z diagonál.

47.2 Korelační koeficient je 0,584. Pozor, mlčky jsme učinili předpoklad, že známky jsou měřeny na metrické (= intervalové nebo poměrové) škále!

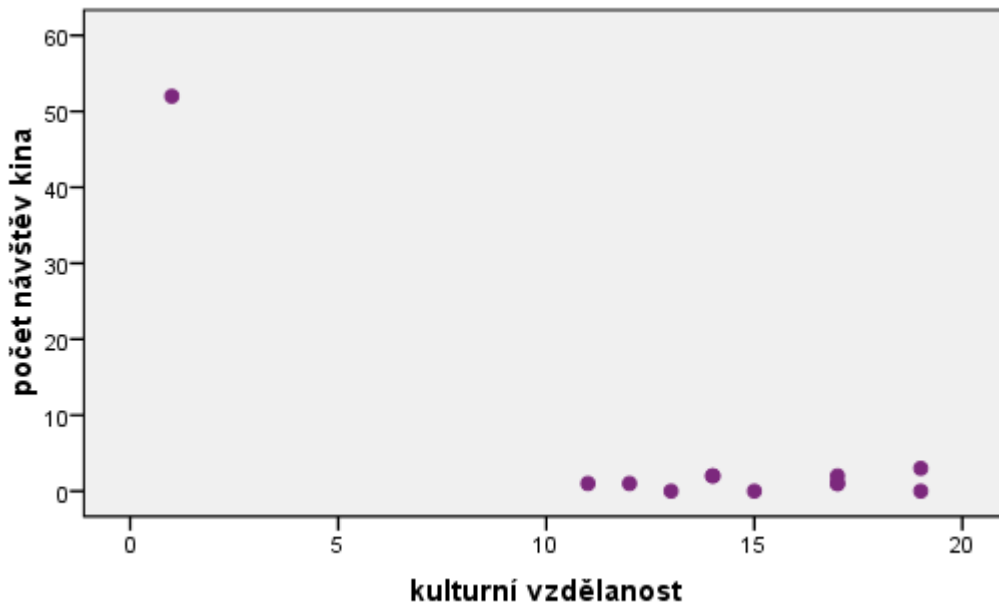
47.3 Nejde o intervalová data, navíc si jsme jistí tím, že intervaly mezi známkami nejsou stejné. Proto je nevhodnější Kendallovo tau. Vzhledem k tomu, že se zde vyskytují stejné hodnoty (tedy i pořadí), je třeba použít tau-b, které se počítá o něco složitěji, než jak jsme si tau představili. Vychází 0,495.

48.1 Pearsonův korelační koeficient je -0,837. Naznačuje tedy velmi výrazný negativní vztah.

48.2 Outlier Karel ovlivňuje velmi výrazně Pearsonův korelační koeficient. Pokud jej vynecháme, bude korelace výrazně nižší. Pearsonův korelační koeficient je momentová statistika, odlehle hodnoty na něj mají velmi velký vliv. Pokud tedy budeme mít ve výzkumu podobně nesmyslná data, je nutné vynechat je. Ovšem často nejsou tato data tak zřejmá jako tady.

Pearsonův korelační koeficient dat bez osoby číslo 9 vychází 0,204, tedy prakticky žádný vztah. A dokonce opačného směru než v předchozím případě.

Následující scatter tato data zobrazuje.



48.3 Byl by Karlem ovlivněn podstatně méně. Zkuste si ho nechat spočítat (nepočítejte ho sami, u návštěv kina jsou shodné hodnoty a s nimi je výpočet složitější).

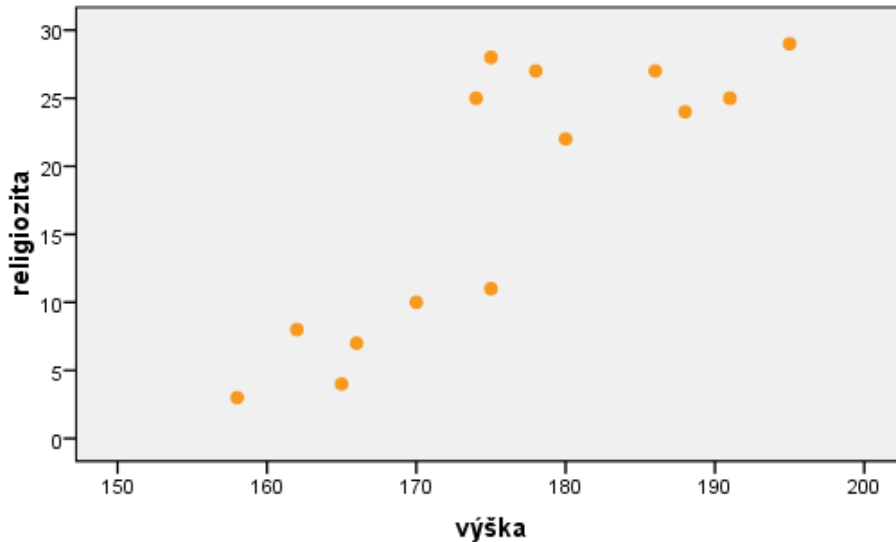
48.4 Také by měl být ovlivněn méně než Pearsonův, avšak více než tau.

48.5 Resistance.

49.1 Pearsonův korelační koeficient bude vycházet vysoký kladný. Je to způsobeno dvěma odlišnými skupinami, přitom jedna jsou vysocí a religiozní, druhá nízcí a nevěřící. Pearsonův korelační koeficient je velice ovlivňován pozicí těchto dvou skupin vůči sobě. Pearsonův korelační koeficient zde vychází 0,838, naznačuje velmi silný pozitivní vztah.

49.2 Pearsonův korelační koeficient matčinych příbuzných (osoby 1 – 8) je 0,172. Pearsonův korelační koeficient otcových příbuzných (9 – 14) je 0,831. Vztah mezi výškou a religiozitou v jednotlivých skupinách by mohl být v podstatě libovolný. Pokud však do výzkumu zahrneme dvě takto výrazně odlišné skupiny, pak za vysokou hodnotu Pearsonova korelačního koeficientu může takřka pouze vyhraněnost skupin.

Scatter opět ukazuje data.



50.1 Pearsonův korelační koeficient je určen pro metrická data, tedy měřena na intervalové či poměrové škále. Pořadí v běhu je pouze na ordinální úrovni, Pearsonův korelační koeficient použít nemůžeme. Je potřeba buď Spearmanův nebo Kendallův koeficient pořadové korelace.

50.2 Tabulka zobrazuje data seřazená podle výšky.

výška v cm	164	168	170	173	177	180	182	185	189	191
pořadí v běhu	7	9	5	10	2	8	4	1	6	3

Počet konkordancí: $P = 14$

Počet diskordancí: $Q = 31$

Kendallovo $S = P - Q = -17$

Kendallův korelační koeficient: $t_k = 2S / n(n-1) = -0,378$

51.

a) Pokud víme o vlivu velikosti města na obě proměnné, neměli bychom jej ignorovat. Vztah mezi K a S je patrně pouze zdánlivý. Ale spočítejme si to...

b) $r_{KS,P} = 0,15$

52.

a) $r_{KP} = 0,5$

b) Školní docházka trénuje schopnost udržet pozornost a zvyšuje slovní zásobu použitelnou pro tvorbu rýmů. Mohla by nám skutečný vztah mezi kreativitou a pozorností nadsazovat....

c) $r_{KP,S} = 0,22$

53.

a) korelace (1-2, 2-3, 1-3) jsou 0,73 0,93 0,78. To jsou kladné (a vysoké) korelace, takže sčítat určitě můžeme.

b) $\alpha = 0,92$ To je příjemně vysoká hodnota.