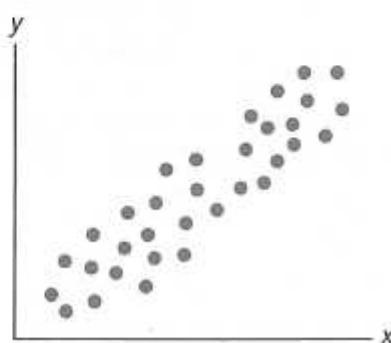
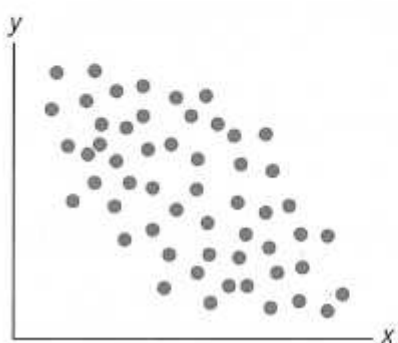
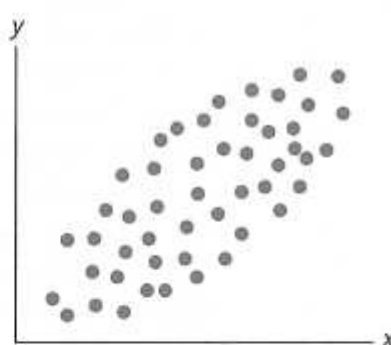
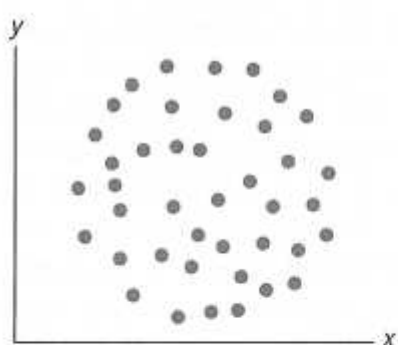


Otázky k lineární korelaci

1. Jaká je odlišnost mezi korelačním a experimentálním designem? Proč nabízí experimentální design možnost formulace kauzálních tvrzení o vztahu mezi dvěma proměnnými?
2. Uveďte příklad proměnných, které korelují pozitivně a proměnných korelujících negativně.
3. Jaké informace můžeme shromáždit na základě bodového grafu (scatter plot)?
4. Je dán soubor z-skórů, jaká je korelace mezi X a Y?

z_x	z_y
-.32	-.56
-.10	0
.42	-.12
0	.68

5. Odhadněte r pro každý z těchto bodových grafů:



6. Jsou dány následující korelace a stupně volnosti. Jaké jsou kritické/mezní r pro 5 procentní a 1 procentní hladinu významnosti? V každém z případů rozhodněte, zda byste přijali nebo zamítli nulovou hypotézu, a to na 5 i 1 procentní hladině (zkuste to v excelu, ať to nemusíte počítat 8x) :

- a. $r = .67, df = 24$
- b. $r = .39, df = 24$
- c. $r = .89, df = 12$
- d. $r = .74, df = 18$
- e. $r = .45, df = 29$
- f. $r = .95, df = 7$
- g. $r = .62, df = 13$
- h. $r = .24, df = 100$

7. Pro následující soubor dat:

- a. Spočítejte r (funkcí v excelu) a proveďte oboustranný test významnosti na hladině $\alpha = 0,05$.
- b. Specifikujte nulovou a alternativní hypotézu.
- c. Jaké je kritické r ?
- d. Přijímáte nebo zamítáte nulovou hypotézu?
- e. Jaké procento rozptylu Y je vysvětleno rozptylem X ?

X	Y
6	4
7	5
7	6
4	6

8. Pro následující soubor dat:

- a. Spočítejte r (funkcí v excelu) a proveďte oboustranný test významnosti na hladině $\alpha = 0,05$.
- b. Specifikujte nulové a alternativní hypotézy.
- c. Jaké je kritické r ?
- d. Přijímáte nebo zamítáte nulovou hypotézu?
- e. Jaké procento rozptylu Y je vysvětleno rozptylem X ?

X	Y
6	3
9	7
7	6
10	9

9. Pro následující soubor dat:

- a. Spočítejte r (funkcí v excelu) a proveďte oboustranný test významnosti na hladině $\alpha = 0,05$.
- b. Specifikujte nulové a alternativní hypotézy.
- c. Jaké je kritické r ?
- d. Přijímáte nebo zamítáte nulovou hypotézu?
- e. Jaké procento rozptylu Y je vysvětleno rozptylem X ?

X	Y
10	12
9	6
11	10
13	13

10. Pro následující soubor dat:

- Spočítejte r (funkcí v excelu) a proveďte oboustranný test významnosti na hladině $\alpha = 0,05$.
- Specifikujte nulové a alternativní hypotézy.
- Jaké je kritické r ?
- Přijímáte nebo zamítáte nulovou hypotézu?
- Jaké procento rozptylu Y je vysvětleno rozptylem X ?

X	Y
1	3
2	4
4	5
6	5
7	7

12. Zodpovězte následující otázky, je-li dán soubor dat:

- Spočítejte r (funkcí v excelu) a proveďte oboustranný test významnosti na hladině $\alpha = 0,05$.
- Specifikujte nulové a alternativní hypotézy.
- Jaké je kritické r ?
- Přijímáte nebo zamítáte nulovou hypotézu?
- Jaké procento rozptylu Y je vysvětleno rozptylem X ?

X	Y
1	8
3	7
3	6
5	5
6	4

13. Nakreslete bodové grafy pro data z otázek č. 7, 8, 9 a 12 (zkuste si to i v Excelu, SPSS, či Statistice).

14. Pro skupinu 75 subjektů se $\sum(z_x z_y) = 64$. Jaké je ρ ?

15. Jakým způsobem ovlivňuje variabilita výběrových skóre velikost korelace?

16. Jakým způsobem ovlivňuje korelaci použití extrémních skupin?

17. Jestliže použijete Pearsonův korelační koeficient na spočítání korelace v případě, kdy je vztah mezi X a Y nelineární, jaký to má vliv na r ?

18. Sportovní psycholog se zajímá o vztah mezi počtem týdnů cvičení a klidovým srdečním tepem. Na základě dat určete:

- Jaké je r (Excel :-)?
- Jaké jsou nulové a alternativní hypotézy?
- Jaké je kritické r pro nepřímý test na 5 procentní hladině významnosti?
- Jaké procento rozptylu klidového srdečního tepu je vysvětleno počtem týdnů cvičení?

e. Měli byste akceptovat nebo zamítnout nulovou hypotézu?

<i>Weeks of Exercise</i>	<i>Resting Heart Rate</i>
2	82
4	78
8	72
14	66
10	66
9	70
9	69

19. Školní psycholog předpokládá vztah mezi IQ a počtem sourozenců. Z uvedených dat odpovězte na otázky:

- Jaké je r (Excel :-)?
- Jaké jsou nulové a alternativní hypotézy?
- Jaké je kritické r pro nepřímý test na 5 procentní hladině významnosti?
- Jaké procento rozptylu klidového srdečního tepu je vysvětleno počtem týdnů cvičení?
- Měli byste akceptovat nebo zamítnout nulovou hypotézu?

<i>Number of Siblings</i>	<i>IQ</i>
8	123
3	100
1	90
4	111
2	102
0	95

20. Psycholog se zajímá o sílu asociace mezi věkem a vykonáním určitých úkolů, které vyžadují motorické dovednosti. Načrtněte scatter pro následující data a rozhodněte o nejvhodnějším postupu pro testování hypotézy:

<i>Age in Years</i>	<i>Number of Errors</i>
6	23
7	19
8	17
9	16
10	16
11	18
12	18
13	19
14	20
15	22

Otázky pro počítačové zpracování:

21. Baron, Logan a Kao (1990) studovali vztah mezi nepohodlím pacientů, které bylo posuzováno studenty stomatology a samotnými pacienty. Nepohodlí bylo definováno jako

kombinace úzkosti, bolesti a distresu (*nižší hodnoty/čísla indikují nižší nepohodlí*). Jednotlivé hodnoty byly získány za dvou podmínek: během vrtání a během umístění pryžové hráze kolem zubu. (Pryžová hráz je gumová pochva okolo zubu, která je připojena na kovový rámec, izoluje zub a tím brání tomu, aby byly úlomky spolknuty. Umístění této hráze vyžaduje zvýšenou pozornost ze strany zubaře než jednoduchá výplň zubu). Korelace mezi nepohodlím posuzovaným studenty stomatologie a samotnými pacienty byly spočítány za obou podmínek. Korelace mezi hodnoceními během vrtání byly významné: $r(41) = +0,52, p < 0,05$. Korelace během umístění hráze mezi hodnocením studenty stomatologie a pacienty byly také významné, ale o hodně menší: $r(41) = +0,21, p < 0,05$. Autoři spekulují, že schopnost postihnout distres pacientů závisí na tom, co dělají... „a že být senzitivní vůči stresu vyžaduje značnou kapacitu pozornosti (s. 151).“ Následující soubor dat je hypotetický. Hodnoty jsou vybrány, takže vaše korelace povedou k závěrům, které jsou konzistentní s autory. Nezapomeňte prověřit scatterly dat.

Discomfort Ratings During Drilling

<i>Dental Students</i>	<i>Patients</i>
8	6
6	9
3	1
1	4
5	5
4	6
8	8
7	6
9	6
2	3
1	1
6	8
4	6
3	3
9	7
7	8
6	9
2	8
5	7
6	6
3	2
1	1
5	7
6	9
8	8
9	6

Discomfort Ratings During Rubber Dam

<i>Dental Students</i>	<i>Patients</i>
8	6
6	9
3	1
1	4
5	5
4	6
8	8
7	6
9	6
2	3
1	1
6	8
4	6
3	3
9	7
7	8
6	9
2	8
5	7
6	6
3	2
1	1
5	7
6	9
8	1
9	4

22. Carrie (1981) zkoumala vztah mezi podáváním zpráv o symptomech během těhotenství a během menstruace a asociací těchto zpráv s obecnou tendencí podávat zprávy/vypovídat o psychologických a fyzikálních symptomech. Kromě jiného zjistila významnou korelaci mezi počtem symptomů zažívaných během menstruace a počtem symptomů, o kterých podávaly zprávy během těhotenství. Následující data jsou hypotetická a povedou vás ke korelacím, které jsou konzistentní s hladinou významnosti, jak ji lze nalézt ve studii Carrie.

Hypothetical Questionnaire Scores

<i>Last Menstruation Symptoms</i>	<i>Last Pregnancy Symptoms</i>
93	87
75	64
34	78
23	55
76	43
34	45
21	20
34	54
60	60
45	82
67	67
50	48
89	72
61	68
56	45
82	75
45	34
53	55
71	50
59	90
90	56
43	62
49	32

Odpovědi

1. V korelačním designu je studie vedena bez možnosti kontroly nad studovaným fenoménem. V experimentálním designu některé procedurální proměnné zůstávají konstantní a jinými se účelově manipuluje. Manipulací NP v experimentálním designu existuje větší pravděpodobnost, že budete schopni ukázat kauzální vztah, který má efekt NP na ZP.

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

$$4. \rho = \frac{.13}{4} = .03$$

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

6.	$r_{\text{crit}} 5\%$	$r_{\text{crit}} 1\%$	$r_{\text{crit}} 5\%$	$r_{\text{crit}} 1\%$
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

(Kritické hodnoty pro h jsou přibližné, založené na $df = 100$)

$$7. \text{ a. } \begin{array}{lll} \Sigma X = 24 & \Sigma Y = 21 & \Sigma XY = 125 \\ (\Sigma X)^2 = 576 & (\Sigma Y)^2 = 441 & n_p = 4 \\ \Sigma X^2 = 150 & \Sigma Y^2 = 113 & \end{array}$$

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

- b. $H_0: \rho = 0; H_1: \rho \neq 0$
c. $r_{\text{crit}}(2) = .950$
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.)

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

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

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

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.)

10. a. $\Sigma X = 20$ $\Sigma Y = 24$ $\Sigma XY = 110$
 $(\Sigma X)^2 = 400$ $(\Sigma Y)^2 = 576$ $n_p = 5$
 $\Sigma X^2 = 106$ $\Sigma Y^2 = 124$

$$r_{\text{obt}} = \frac{70}{75.63} = .93$$

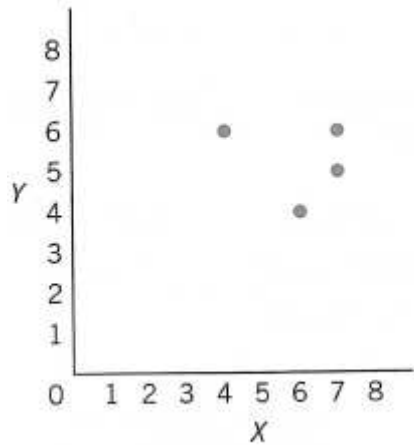
- b. $H_0: \rho = 0; H_1: \rho \neq 0$
c. $r_{\text{crit}}(3) = .878$
d. Reject the null hypothesis.
e. 86.5 percent

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

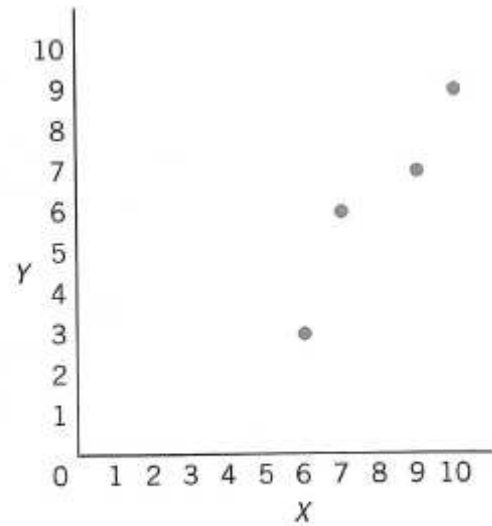
$$r_{\text{obt}} = \frac{-60}{61.64} = -.97$$

- 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.

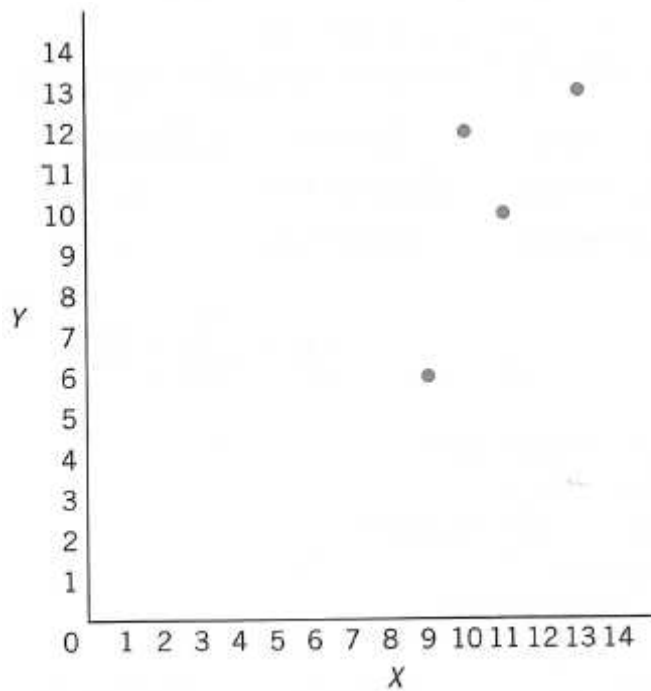
13. Grafy:



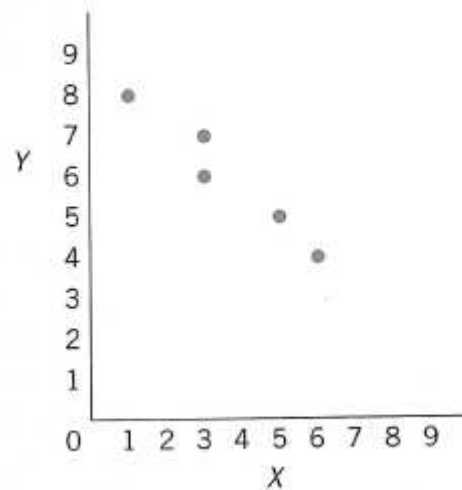
(Problem 7)



(Problem 8)



(Problem 9)



(Problem 12)

14. $\rho = \frac{64}{75} = .85$

15. Omezený rozsah skóre může podhodnotit velikost korelace v populaci.

16. Použití extrémních skupin může zvýšit velikost korelace.

17. r bude podhodnocovat ρ .

$$\begin{array}{lll} 18. \text{ a. } & \Sigma X = 56 & \Sigma Y = 503 & \Sigma XY = 3887 \\ & (\Sigma X)^2 = 3136 & (\Sigma Y)^2 = 253,009 & n_p = 7 \\ & \Sigma X^2 = 542 & \Sigma Y^2 = 36,365 & \end{array}$$

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

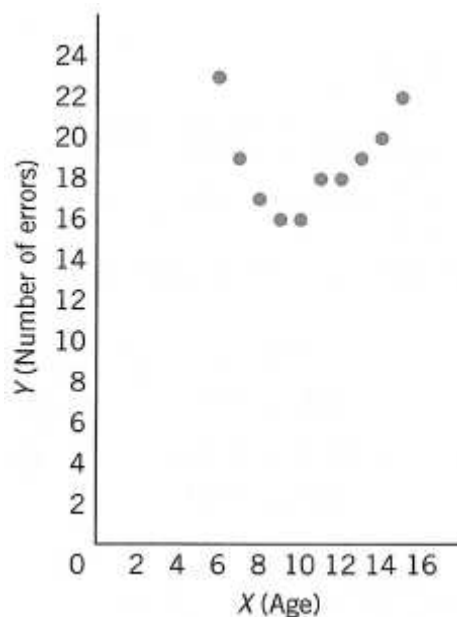
- b. $H_0: \rho = 0; H_1: \rho \neq 0$
- c. $r_{\text{crit}}(5) = .754$
- d. 90.25 percent
- e. Reject the null hypothesis.

$$\begin{array}{lll} 19. \text{ a. } & \Sigma X = 18 & \Sigma Y = 621 & \Sigma XY = 2022 \\ & (\Sigma X)^2 = 324 & (\Sigma Y)^2 = 385,641 & n_p = 6 \\ & \Sigma X^2 = 94 & \Sigma Y^2 = 64,979 & \end{array}$$

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

- 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.

20. Graf:



21. Drilling: $r(24) = .65, p < .05$
Rubber Dam: $r(24) = .46, p < .05$
22. $r(21) = .413, p < .05$

Otázky k lineární regresi

1. Jaké jsou dvě důležitá použití regresní analýzy?
2. Co znamená, že regresní přímka je založená na metodě nejmenších čtverců?
3. Proč nemůže být $\sum(Y - Y')$ použita jako ukazatel chyby odhadu?
4. Jestliže $r = 0$, čemu musí být b rovno?
5. Popište, jakým způsobem se r vztahuje k predikci/odhadu chyby.
6. Když použijete směrodatnou odchylku reziduálních hodnot jako ukazatel chyby predikce, proč je důležitý předpoklad homoskedasticity?
7. Proč musí být výzkumník opatrný, použije-li rovnici predikce u jedince, které má odlišné charakteristiky od lidí, kteří byli zahrnuti v originálním vzorku, z kterého byla rovnice regrese odvozena?
8. Ve studiu tolerance k bolesti se výzkumník zajímá o predikci/odhad doby, během které jsou subjekty schopni udržet ruce ve velmi ledové vodě. Z dřívějšího výzkumu ví, že příjem vitamínu E během minulých 12 hodin koreluje s tolerancí bolestivého stimulu, přinejmenším, pokud je jím zmrzlá voda. Následující tabulka ukazuje páry skóre pro studovaný vzorek:

Vitamin E: (X)	Tolerance Times (in seconds): (Y)
5	23
9	32
22	65
12	40
16	42

- a. Jaká je směrnice b (potřebné statistiky si nechte spočítat v excelu)?
 - b. Jaký je průsečík a ?
 - c. Jaké je směrodatná odchylka chyb odhadu s_e ?
 - d. Jaký čas tolerance byste predikovali někomu, kdo si vzal ráno při studii 16 jednotek vitamínu E?
9. Sociolog se zajímá o predikci ročního platu (Y) založenou na dřívější úrovni vzdělání (X). Úroveň vzdělání je definována jako počet let školní docházky. Následující data byla získána od 6 subjektů:
- a. Jaká je směrnice b (potřebné statistiky si nechte spočítat v excelu)?
 - b. Jaký je průsečík a ?
 - c. Jaké je směrodatná odchylka chyb odhadu s_e ?
 - d. Jaký příjem byste odhadovali pro někoho s desetiletou docházkou?

<i>Education: (X)</i>	<i>Income X 1000: (Y)</i>
10	15
14	29
9	14
14	37
12	20
13	23

10. Vstupní komise potřebuje odhadnout, zda určitý student bude schopný získat potřebné známky během prvního roku na vysoké škole. Aby prošel prvním rokem, musí dosáhnout 3,00 GPA (průměrný prospěch, vyšší čísla lepší) v prvním semestru na univerzitě. Vstupní komise má data z dřívějších let, a to GPA ze střední školy a GPA z prvního roku na univerzitě. Data jsou následující:

<i>Undergraduate GPA: (X)</i>	<i>Graduate School GPA: (Y)</i>
3.50	3.33
3.98	3.63
3.10	3.40
2.90	3.41
3.40	3.40

- Jaké je b ?
- Můžete přijmout studenta se středoškolským GPA = 3,00?
- Mezi jakými hodnotami výsledků GPA v prvním roce na univerzitě by se mělo ocitnout 68% studentů, kteří vstupovali do studia se středoškolským GPA = 3,67-

Otázky pro počítačové zpracování

11. Na základě otázky č. 21 k lineární korelaci, vypočítejte rovnici regrese a s_e pro obě podmínky. Jestliže během vrtání student stomatologie ohodnotí pacientovo nepohodlí 7, jak byste odhadli/predikovali hodnocení nepohodlí samotným pacientem? Odpovězte na stejné otázky i pro druhou podmínku a určete standardní chybu odhadu.

12. Na základě otázky č. 22 k lineární korelaci, pro ženu, která vypovídá o 54 menstruačních symptomech, určete, o kolika symptomech bude vypovídat během těhotenství? Identifikujte dvě hodnoty těhotenských symptomů, mezi kterými se nachází 68 procent žen, které vypovídají o 54 menstruačních symptomech.

Odpovědi

1. Regresní analýza je důležitá pro predikce týkající se praktických problémů a při budování a testování teorií.
2. Regresní přímka je k bodovému grafu v takovém vztahu, že suma druhých mocnin odchylek jednotlivých bodů od regresní přímky ($\sum e^2$) je minimalizována.
3. $\sum(Y - Y') = 0$.
4. Jestliže $r = 0$, pak $b = 0$.
5. Se vzrůstající korelací, chyba odhadu klesá.
6. Předpoklad homoskedasticity nám umožňuje předpokládat, že s_e je stejné pro každou hodnotu X. Jestliže by byla každá úroveň X spojena s jinak distribuovanými chybami, musela by být pro každý skóre X odlišná chyba odhadu.
7. Generalizace na populaci odlišné od té, z které jsme získali regresní rovnici, může vést k neplatným predikcím..

8. $\bar{X} = 12.80$ $\bar{Y} = 40.40$
 $\Sigma X = 64$ $\Sigma Y = 202$ $\Sigma XY = 2985$
 $(\Sigma X)^2 = 4096$ $(\Sigma Y)^2 = 40,804$ $n_p = 5$
 $\Sigma X^2 = 990$ $\Sigma Y^2 = 9142$

a. $b = \frac{1997}{854} = 2.34$

b. $Y_p = 40.40 + 2.34(0 - 12.80) = 10.45$

c. $s_e = \sqrt{\left[\frac{1}{5(3)} \right] \left[5(9142) - 40,804 - \left(\frac{[5(2985) - (64)(202)]^2}{5(990) - 4096} \right) \right]} = 4.07$

d. $Y_p = 40.40 + 2.34(16 - 12.80) = 47.89$ seconds

9. $\bar{X} = 12$ $\bar{Y} = 23$
 $\Sigma X = 72$ $\Sigma Y = 138$ $\Sigma XY = 1739$
 $(\Sigma X)^2 = 5184$ $(\Sigma Y)^2 = 19,044$ $n_p = 6$
 $\Sigma X^2 = 886$ $\Sigma Y^2 = 3560$

a. $b = \frac{10,434 - 9936}{5316 - 5184} = 3.77$

b. $Y_p = 23 + 3.77(0 - 12) = - 22.24$

$$c. s_e = \sqrt{\left[\frac{1}{6(4)} \right] \left[6(3560) - 19,044 \right] - \left(\frac{[6(1739) - (72)(138)]^2}{6(886) - 5184} \right)} = 4.18$$

$$d. Y_p = 23 + 3.77(10 - 12) = 15.46 (\$15,460)$$

$$10. \quad \begin{array}{lll} \bar{X} = 3.38 & \bar{Y} = 3.43 & \\ \Sigma X = 16.88 & \Sigma Y = 17.17 & \Sigma XY = 58.09 \\ (\Sigma X)^2 = 284.93 & (\Sigma Y)^2 = 294.81 & n_p = 5 \\ \Sigma X^2 = 57.67 & \Sigma Y^2 = 59.01 & \end{array}$$

$$a. b = \frac{290.45 - 289.83}{288.35 - 284.93} = .18$$

$$b. Y_p = 3.43 + .18(3.00 - 3.38) = 3.36$$

Yes, since we would predict the student to achieve a GPA of 3.36 (3.00 minimum required).

$$c. s_e = .095$$

$$Y_p = 3.43 + .18(3.67 - 3.38) = 3.48$$

$$3.48 \pm .095 = 3.575 \text{ and } 3.385$$

$$11. \text{ Drilling: } Y_p = 5.77 + .64(X - 5.15)$$

$s_e = 1.98$ (A small discrepancy is due to rounding error.)

$$Y_p = 5.77 + .64(7 - 5.15) = 6.95$$

$$\text{Rubber Dam: } Y_p = 5.42 + .47(X - 5.15)$$

$s_e = 2.44$ (A small discrepancy is due to rounding error.)

$$Y_p = 5.42 + .47(7 - 5.15) = 6.29$$

$$12. Y_p = 58.35 + .35(54 - 56.96) = 57.31$$

$$s_e = 16.36$$

$$Y_p = 57.31 \pm 16.36 = 73.67 \text{ and } 40.95$$