Odpovědi (téma 6)

1.1 a

1.2 b – korelace je součástí regrese a je bez rozměrů

1.3 jde o transformaci hodnot proměnných, jejichž regresní vztah není lineární (nelze jej popsat přímkou, nýbrž křivkou); transformace má za cíl, aby mezi novými proměnnými existoval lineární vztah, který bude možno analyzovat lineární regresní analýzou

2.1 Regresní analýza je důležitá pro predikce týkající se praktických problémů a při budování a testování teorií.

2.2 Předpoklad homoskedascity nám umožňuje předpokládat, že s_e je stejné pro každou hodnotu X. Jestliže by bylo každá úroveň X spojena s jinak distribuovanými chybami, musela by být pro každý skór X odlišná chyba odhadu.

2.3 Regresní přímka je k bodovému grafu v takovém vztahu, že suma druhých mocnin odchylek jednotlivých bodů od regresní přímky na škále predikované proměnné ($\sum e^2$) je minimalizována.

2.4 $\sum(Y - Y') = 0$

2.5 Se vzrůstající korelací chyba odhadu klesá.

2.6 ne, klesá (tato otázka se v zásadě jinými slovy ptá na totéž co otázka předchozí)

2.7 Generalizace na populaci odlišnou od té, z které jsme získali regresní rovnici, může vést k neplatným predikcím

3.1 $b = 0$.

3.2 ne nutně, ale při standardizovaných skórech ano (tj. $z_X = z_Y$).

3.3 $z_Y' = 1{,}0$

3.4 P_{84}

3.5 $z_Y' = 0{,}9$

3.6 ano

3.7 $s_e = 8$

3.8 c

3.9 u jednoduché lineární regrese platí, že $r^2 = s_{reg}^2 / s_Y^2$, a proto $s_{reg}^2 = r^2 * s_Y^2 = 0{,}6^2 * 10^2 = 36$

3.10 $s_Y^2 = s_{reg}^2 + s_{res}^2$, a proto $s_{res} = \sqrt{(s_Y^2 - s_{reg}^2)} = \sqrt{(100 - 36)} = 8$

4.1 68%

4.2 16%

4.3 ano

5.1 115

5.2 95

5.3 100

5.4 15 x (0,8) = 12

5.5 32%

6.1 – 6.4

8. $\overline{X} = 12.80$ $\overline{Y} = 40.40$

$\Sigma X = 64$ $\Sigma Y = 202$ $\Sigma XY = 2985$

$(\Sigma X)^2 = 4096$ $(\Sigma Y)^2 = 40{,}804$ $n_p = 5$

$\Sigma X^2 = 990$ $\Sigma Y^2 = 9142$

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

Zde rozepsané řešení (Excelový objekt, dvojklikem se dostanete ke vzorečkům)

X	Y	Y'	res	res2
5	23	22,16	0,84	0,70
9	32	31,51	0,49	0,24
22	65	61,91	3,09	9,53
12	40	38,53	1,47	2,16
16	42	47,88	-5,88	34,61
12,800	40,400	m	0,000	15,75
6,535	15,662	sd	3,437	3,97
	0,976	r		sigma e
	2,338	b		
	10,468	a		dle komput
	3,437	se (pomocí r2)		3,968209

7.1 – 7.4

9. $\overline{X} = 12$ $\overline{Y} = 23$

$\Sigma X = 72$ $\Sigma Y = 138$ $\Sigma XY = 1739$

$(\Sigma X)^2 = 5184$ $(\Sigma Y)^2 = 19{,}044$ $n_p = 6$

$\Sigma X^2 = 886$ $\Sigma Y^2 = 3560$

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

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

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

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

8.1 – 8.3 a 9.1 – 9.2

X	Y	Y'	res
3,5	3,33	3,46	-0,13
3,98	3,63	3,54	0,09
3,1	3,4	3,38	0,02
2,9	3,41	3,35	0,06
3,4	3,4	3,44	-0,04
3,376	3,434	*M*	0,000
0,413	0,114	*SD*	0,085
	0,665	*r*	
	0,184	*b*	
	2,814	*a*	
	0,085	s_e	

10. $\overline{X} = 3.38$ $\overline{Y} = 3.43$

$\Sigma X = 16.88$ $\Sigma Y = 17.17$ $\Sigma XY = 58.09$

$(\Sigma X)^2 = 284.93$ $(\Sigma Y)^2 = 294.81$ $n_p = 5$

$\Sigma X^2 = 57.67$ $\Sigma Y^2 = 59.01$

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

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. $s_e = .095$
$Y_p = 3.43 + .18(3.67 - 3.38) = 3.48$
$3.48 \pm .095 = 3.575$ and 3.385

11. Drilling: $Y_p = 5.77 + .64(X - 5.15)$
$s_e = 1.98$ (A small discrepancy is due to rounding error.)

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

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. $Y_p = 58.35 + .35(54 - 56.96) = 57.31$
$s_e = 16.36$
$Y_p = 57.31 \pm 16.36 = 73.67$ and 40.95

10.1 64 a 146

10.2 55 a 138

10.3 ano

10.4 ano

10.5 b = 0,694

10.6 a = 30,5

10.7 Y' = 0,694X + 30,5

10.8 128

10.9 79

10.10 -

10.11 s_e = 6,9

10.12 cca 68%

10.13 Tomáš mezi 121 a 135 (128 +/- s_e); David mezi 72 a 86

11.1 Y' = 0,11X' – 3

11.2 m = 8

11.3 m = 6,9

11.4 predikovaný skór má vyšší percentilový ekvivalent (P_{29}) než hodnota prediktoru (P_{25})

11.5 s_e = 1,1 a predikovaný skór pro IQ = 90 je 6,9. $m_č$ = 8, takže chyby o více než jednu s_e směrem nahoru budou nad 8. Nad z=1 je 16% rozložení. Tj. cca 16%.

12.1 $b = r * (s_y / s_x)$, a proto $r = b / (s_y / s_x) = 0,5 / (2,0 / 0,8) = 0,2$
$R^2 = r^2 = 0,2^2 = 0,04$, což jsou 4 %
depresivitou lze tedy vysvětlit 4 % rozptylu chatování

12.2 $b = r * (s_y / s_x) = 0,2 * (0,8 / 2,0) = 0,08$
$a = m_y – b*m_x = 1,6 – 0,08*0,6 = 1,55$
regresní rovnice je tudíž y = 0,08x + 1,55

12.3 y = 0,08 * 10 + 1,55 = 2,35

13.1 MG = 72 – 4BC = 72 – 4*1 = 68

13.2 $r = b / (s_y / s_x) = 4 / (10/1,5) = 0,6$
$s_{reg}^2 = s_y^2 * r^2 = 10^2 * 0,6^2 = 36$
$s_Y^2 = s_{reg}^2 + s_{res}^2$, a proto $s_{res} = \sqrt{(s_Y^2 – s_{reg}^2)} = \sqrt{(100 – 36)} = 8$
nebo
$s_{res} = \sqrt{(s_Y^2 * (1-r^2))} = \sqrt{(100 * 0,64)} = 8$
predikovaný skór uchazeče je 68 (viz předchozí podotázku); jak jsme právě spočítali, směrodatná odchylka reziduálních hodnot je 8; proto pokud by bylo uchazečovo výsledné skóre menší než šedesát, znamenalo by to chybu odhadu (reziduál) s hodnotou menší než 1s; vzhledem k tomu, že chyby odhadu musejí být rozloženy normálně, pravděpodobnost výskytu takovéto chyby je 16 %; pravděpodobnost nepřijetí uchazeče je proto 16 %

14.1 bodový graf (scatterplot)

14.2 $R^2 = r^2 = 0,49$

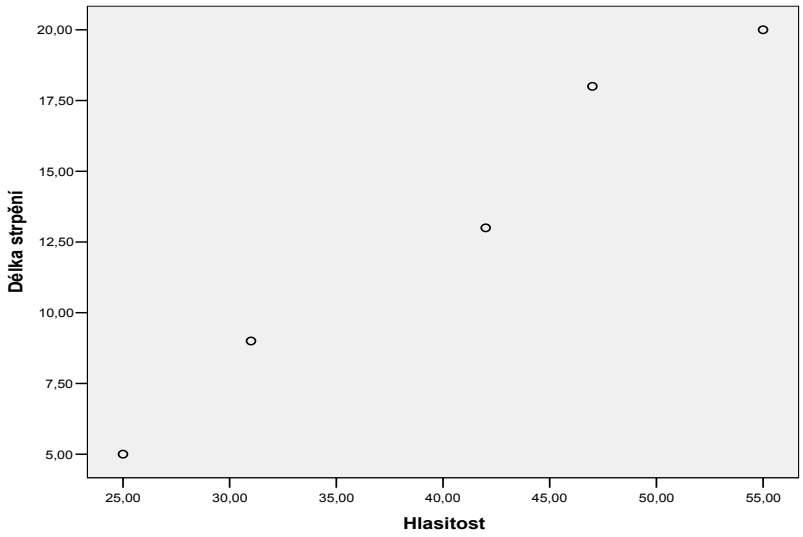
14.3 $d = 2r / \sqrt{(1 – r^2)} = -1,96$

14.4 $s_{reg}^2 = s_y^2 * r^2 = 8^2 * (-0,7)^2 = 31,36$
$s_{res}^2 = s_Y^2 – s_{reg}^2 = 64 – 31,36 = 32,64$
nebo
$s_{res} = \sqrt{(s_Y^2 * (1-r^2))} = \sqrt{(64 * 0,51)} = \sqrt{32,64}$

14.5 příliš ne, protože regresní vztah by byl popsán lépe křivkou než přímkou

14.6 ano; je to i pěkně vidět na grafu – spokojenosti o hodnotě 0,00 odpovídá depresivita o hodnotě okolo 12, zatímco spokojenosti o hodnotě 0,10 depresivita o hodnotě 5, čili zhruba o 7 bodů nižší.

15.1



15.2 oba dva koeficienty budou mít hodnotu 1

15.3 $b = r * (s_y / s_x) = 0,98 * (6/12) = 0,49$
$a = m_y - b*m_x = 13 – 0,49*40 = -6,6$
regresní rovnice je tedy: $y' = 0,49x – 6,6$

15.4 $s_{reg}^2 = s_y^2 * r^2 = 6^2 * 0,98^2 = 34,57$
$s_Y^2 = s_{reg}^2 + s_{res}^2$, a proto $s_{res} = \sqrt{(s_Y^2 – s_{reg}^2)} = \sqrt{(36 – 34,57)} = 1,2$
nebo
$s_{res} = \sqrt{(s_Y^2 * (1-r^2))} = \sqrt{(36 * 0,04)} = 1,2$

15.5 $y = 0,49x – 6,6 = 22,8$

16.1 lineární vztah – přibližně ano
homoskedascita – přibližně ano
normální rozložení reziduí – nevíme, dokud regresi neprovedeme

16.2 $b = r * (s_y / s_x) = 0,5 * (0,31/0,33) = 0,47$
$a = m_y - b*m_x = 1,5 – 0,47*1,6 = 0,75$
regresní rovnice tedy je $y' = 0,47x + 0,75$

16.3 je třeba na základě regresní rovnice spočítat dva libovolné body (např. pro x = 1 [1; 1,22] a x = 2 [2; 1,69]) a ty spojit přímkou

16.4 $R^2 = r^2 = 0,5^2 = 0,25$, tj. 25 %

16.5 $s_{reg}^2 = s_y^2 * r^2 = 0,31^2 * 0,5^2 = 0,024$; $s_{reg} = 0,155$

16.6 průměr je z definice 0; $s_{res} = \sqrt{(s_Y^2 – s_{reg}^2)} = \sqrt{(0,31^2 – 0,024)} = 0,27$

17.1 $r = b / (s_y / s_x) = 0,71 / (0,31/0,28) = 0,64$

17.2 průměr je z definice 0
$s_{reg}^2 = s_y^2 * r^2 = 0,31^2 * 0,64^2 = 0,039$
$s_Y^2 = s_{reg}^2 + s_{res}^2$, a proto $s_{res} = \sqrt{(s_Y^2 – s_{reg}^2)} = \sqrt{(0,096 – 0,039)} = 0,24$
nebo
$s_{res} = \sqrt{(s_Y^2 * (1-r^2))} = \sqrt{(0,096 * 0,59)} = 0,24$

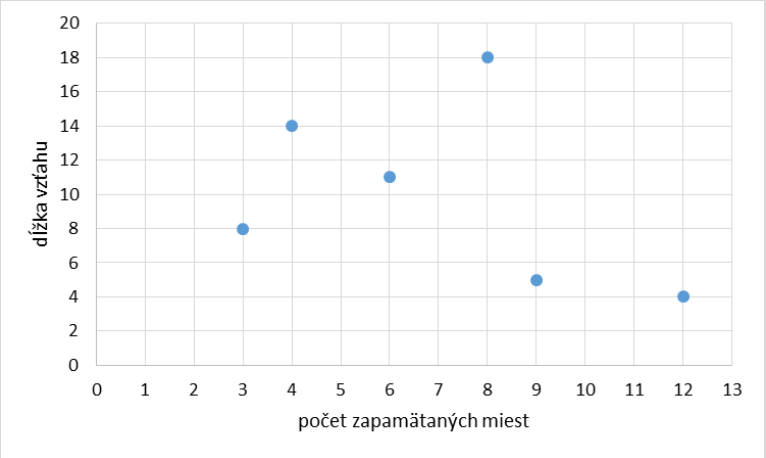
17.3 na základě předchozí podotázky víme, že hodnota 0,24 odpovídá jedné směrodatné odchylce reziduálních skórů; vzhledem k normálnímu rozložení reziduálních skórů předpokládáme, že hodnoty 0 +/-

1s nabude 68 % z nich; pravděpodobnost, že chyba odhadu bude mít velikost -0,24 až 0,24 body je tedy 68 %

18. Najskôr si zosumarizujeme dáta, ktoré potrebujeme k riešeniu úlohy.

	Počet zapamätaných miest	Dĺžka vzťahu
Janů	12	4
Petrů	*9*	5
Vojtů	3	8
Jirků	6	11
Mirků	4	14
Nechoďdomů	8	18

18.1 Údaje sme zobrazili v scatterplote. I keď je údajov menej, je možné si všimnúť, že so stúpajúcim počtom zapamätaných miest klesá dĺžka vzťahu, čiže vzťah medzi oboma premennými bude negatívny približne -0,4.



18.2 Pre predikovanie dĺžky vzťahu v rodine kde si manželka pamätá len 2 miesta si potrebujeme vyrátať lineárnu rovnicu, kde Y=a+bx
$b=r_{dĺžka,miesta}\ (s_{dĺžka}/s_{miesta})$
$a=m_{dĺžka}$ - bm_{miesta}
doplníme si potrebné údaje a to korelačný koeficient, aritmetické priemery a štandardné odchýlky

	Počet zapamätaných miest	Dĺžka vzťahu
Janů	12	4
Petrů	*9*	5
Vojtů	3	8
Jirků	6	11
Mirků	4	14
Nechoďdomů	8	18
m	**7**	**10**
s	**3,35**	**5,4**

Korelačný koeficient **r=-0,409** (vyšiel tak, ako sme vyčítali z grafu)

- Po dosadení: b=-0,409(3,35/5,4)=-0,25
- a=10-(-0,25*7)=11,75
- rovnica potom vyzerá: Y=11,75-0,25X
- ak si manželka pamätá len 2 miesta, potom dĺžka vzťahu Y=11,75-0,5=**11,25**

18.3 Priemer reziduálnych hodnôt je rovný nule
$s_{res}^2=5,4^2(1-(-0,409^2))=24,28$ (rozptyl)
$s_{res}=4,9$; po zaokrúhlení 5 (štandardná odchylka)

18.4 odpoveď: 32%

19.1 odpoveď: 35
19.2 Cca 60%
19.3 Prostredníctvom zobrazenia histogramu reziduálnych hodnôt, alebo aj bodovým grafom (scatterplotom), ktorý zachytáva vzťah reziduálnych hodnôt k hodnotám nezávislej premennej x
19.4 S každým nárastom vizuálnej pamäte o jednotku narastie odhad kreativity o 1 bod.

20.
- snažíme sa transformovať premenné tak, aby bol vzťah lineárny
- delíme vzorku na podskupiny, v ktorých vzťah za lineárny je možné považovať

21. odpoveď d – extrapolaci

22. To znamená, že když věk stoupne o 10 jednotek, budeme člověku odhadovat o 2,2 jednotky vyšší toleranci.

23. Ak nie je splnená homoskedascita, potom so zvyšujúcou sa hodnotou prediktoru (X) bude narastať aj chyba nášho odhadu premennej Y.

24. Matematický spôsob pre nájdenie funkcie (priamky) ktorá najlepšie popisuje dáta, v regresnej analýze – ktorá najlepšie popisuje predikciu premennej Y z premennej X

25. na základe histogramu reziduí, ak rozloženie reziduí nezodpovedá normálnemu rozloženiu.

26. odpoveď 10

27. odpoveď c

28. áno, je to pravda, nakoľko jeden z bodov, ktorými regresná priamka prechádza, je aritmetický priemer X a Y.

29. odpoveď b

30. odpoveď b

31. priesečník a = -27,3 a smernica b = 2,2