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

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

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

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

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

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

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

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

7.1 – 7.4

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

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

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

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

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

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

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

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

8.1 – 8.3 a 9.1 – 9.2

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

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

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

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

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 nepříjetí uchazeče je proto 16 %

14.1 bodový graf (scatterplot)

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

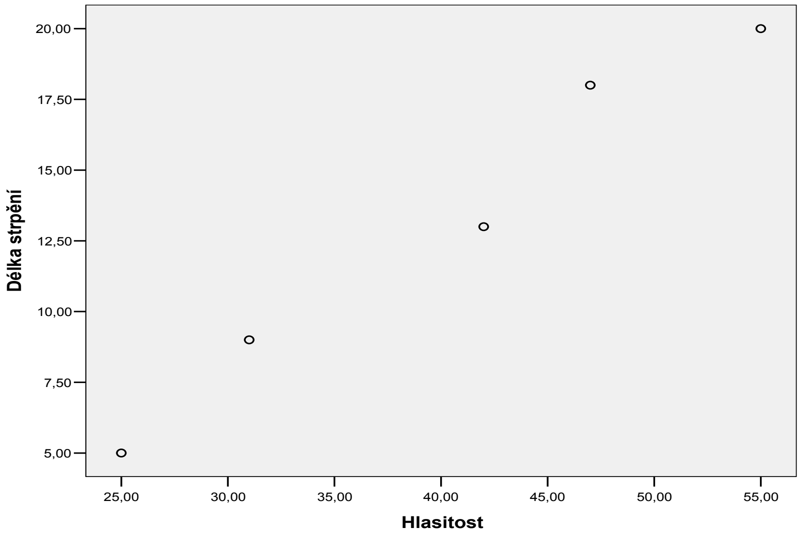
14.3 $d = 2r / \sqrt{(1 - r^2)} = -1{,}96$

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 %