Odpovědi (téma 6)

1.1 a

1.2 b – correlation is part of regression and it is standardized – without the original units

2.1 Regression analysis is important for prediction of practical problems and for building and testing models and theories.

2.2 Homoscedascity is assumption that s_e (residual standard deviation) is the same for every value of variable X.

2.3 Sum of squares of residuals (deviations) of individual values from the regression line ($\sum e^2$) on the scale of predicted variable (Y) is minimalized.

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

2.5 With increasing correlation estimation error decreases.

2.6 no, decreases (the same as in the previous question, just in other words)

2.7 Generalization of results on population different from the population from which we obtained the model can lead to invalid predictions.

3.1 $b = 0$.

3.2 not necessarily, but in standardized scores yes (that is $z_X = z_Y$).

3.3 $z_Y' = 1.0$

3.4 P_{84}

3.5 $z_Y' = 0.9$

3.6 yes

3.7 $s_e = 8$

3.8 c

3.9 in simple linear regression it applies that $r^2 = s_{reg}^2 / s_Y^2$, a proto $s_{reg}^2 = r^2 * s_Y^2 = 0{,}6^2 * 10^2 = 36$

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

4.1 68%

4.2 16%

4.3 yes

5.1 115

5.2 95

5.3 100

5.4 $Se^2 = Sy^2 * (1 - R^2)$, $Se = \sqrt{}(Sy^2 * (1 - R^2))$
$Se = \sqrt{}(15^2 * (1 - 0.6^2) = SQRT(15^2 * (0.64)) = 15 * 0.8 = 12$

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

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

Solution in Excel, by double-click you can see the functions in Excel

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	

9.2

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 and 146

10.2 55 and 138

10.3 yes

10.4 yes

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 Thomas between 121 and 135 (128 +/- s_e); David between 72 a 86

11.1 Y’ = 0.11*X – 3

11.2 m = 8

11.3 m = 6.9

11.4 predicted score has higher percentile equivalent (P_{29}) than predictor (P_{25})

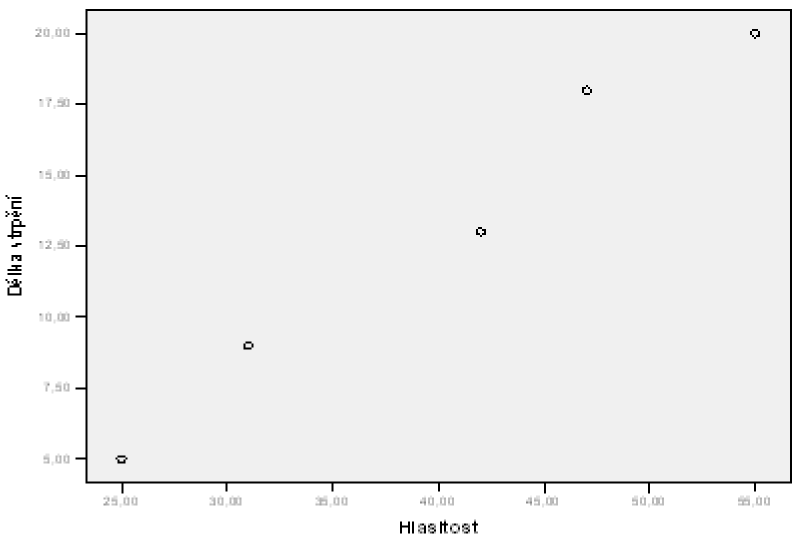
11.5 s_e = 1.1 and predicted score for IQ = 90 is 6.9. *Mr* = 8, thus scores higher than by one s_e will be above 8 (6.9 + 1.1). Above z=1 lie 16% of the distribution, thus cca 16%.

12.1 b = r * (s_y / s_x), thus r = b / (s_y / s_x) = 0.5 / (2.0 / 0.8) = 0.2
$R^2 = r^2 = 0.2^2$ = 0.04, that is 4 %
by depression we can explain 4 % of chatting variance

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
regression equation is then: y = 0.08x + 1.55

12.3 y = 0.08 * 10 + 1.55 = 2.35

15.1



15.2 both coefficients equal 1

15.3 $b = r * (s_y / s_x) = 0.98 * (6/12) = 0.49$
$a = m_y - b*m_x = 13 – 0.49*40 = -6.6$
regression equation is then: $y' = 0.49x – 6.6$

15.4 $s_{reg}^2 = s_y^2 * r^2 = 6^2 * 0.98^2 = 34.57$
$s_Y^2 = s_{reg}^2 + s_{res}^2$, thus $s_{res} = \sqrt{(s_Y^2 – s_{reg}^2)} = \sqrt{(36 – 34.57)} = 1.2$
or
$s_{res} = \sqrt{(s_Y^2 * (1-r^2))} = \sqrt{(36 * 0.04)} = 1.2$

15.5 $y = 0.49x – 6.6 = 22.8$

19.1 35

19.2 Cca 60%

19.3 By displaying histogram of residuals (residual values of estimates), which should have normal distribution and M=0.

20.
- trying to transform variables, so that the relationship is linear
- dividing sample to subsamples in which the relationship is linear

22. It means that if age increases by 10 units, we will estimate by 2.2 units higher tolerance.

23. If homoscedascity is violated, then with increasing value of predictor X will also our estimate error increase.

28. Yes – on of the point through which the regression line goes, is the intercept of Mx and My.

29. b

30. b