Задача 1

1

$$\bar{x} = \frac{0.52 + 0.57 + 0.69 + 0.77 + 0.9 + 0.97 + 1.04 + 1.08 + 1.49 + 1.63}{10} = 0.966$$

$$\bar{y} = \frac{0.28 + 0.33 + 0.34 + 0.34 + 0.33 + 0.38 + 0.46 + 0.49 + 0.52 + 0.49}{10} = 0.396$$

$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = (0.52 - 0.966)(0.28 - 0.396) + (0.57 - 0.966)(0.33 - 0.396) + (0.69 - 0.966)(0.34 - 0.396) + (0.77 - 0.966)(0.34 - 0.396) + (0.9 - 0.966)(0.33 - 0.396) + (0.97 - 0.966)(0.38 - 0.396) + (1.04 - 0.966)(0.46 - 0.396) + (1.08 - 0.966)(0.49 - 0.396) + (1.49 - 0.966)(0.52 - 0.396) + (1.63 - 0.966)(0.49 - 0.396) = 0.25144$

$\sum_{i=1}^{n}(x_i - \bar{x})^2 = (0.52-0.966)^2 + (0.57-0.966)^2 + (0.69-0.966)^2 + (0.77-0.966)^2 + (0.9-0.966)^2 + (0.97-0.966)^2 + (1.04-0.966)^2 + (1.08-0.966)^2 + (1.49-0.966)^2 + (1.63-0.966)^2 = 1.20864$

$\sum_{i=1}^{n}(y_i - \bar{y})^2 = (0.28 - 0.396)^2 + (0.33 - 0.396)^2 + (0.34 - 0.396)^2 + (0.34 - 0.396)^2 + (0.33 - 0.396)^2 + (0.38 - 0.396)^2 + (0.46 - 0.396)^2 + (0.49 - 0.396)^2 + (0.52 - 0.396)^2 + (0.49 - 0.396)^2 = 0.06584$

$$\hat{\rho} = \frac{0.25144}{\sqrt{0.32844 * 2.39184}} = 0.891335030907108$$

$\hat{\rho} > 0$, следовательно, между переменными положительная взаимосвязь.

$0.7 < \hat{\rho} < 0.9$, следовательно, по шкале Чеддока между переменными корреляция высокая.

2

$R^2 = \hat{\rho}^2 = 0.891335030907108^2 = 0.794478137322176$

79,4 % дисперсии результативного признака y объясняется влиянием факторного признака x.

$$y_i = \beta_1 + \beta_2 x_i + e_i$$

$$\hat{\beta}_2 = \frac{Cov(x, y)}{Var(x)} = \frac{0.25144}{1.20864} = 0.208035477892507$$

$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2\bar{x} = 0.396 - 0.208035477892507 * 0.966 = 0.195037728355838$

$$y_i = 0.195 + 0.208x_i + e_i$$

В среднем с ростом з/п на 1 тыс. руб. прожиточный минимум на душу населения в увеличивается на 208 руб.

4

$$F_{obs} = \frac{R^2}{1-R^2}(n-2) = \frac{0.794478137322176}{1-0.794478137322176}(10-2) = 30.9252992152021$$
$$F_{crit} = F_{1,8,0.05} = 5.31765507157871$$

$F_{obs} > F_{crit}$ коэффициент детерминации является статистически значимым, следовательно, рассматриваемая регрессия достаточно определена включенными переменными.

5

$$s_e^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2} = (-0.0232161768599419^2 + 0.016382049245433^2 + 0.00141779189833205^2 - 0.0152250463330685^2 - 0.0522696584590945^2 - 0.01683214191157^2 + 0.0486053746359545^2 + 0.0702839555202541^2 + 0.0149894095843262^2 - 0.0441355573206249^2) : 8 = 0.0016914449298385$$
$$se_{\hat{\beta}_1} = \frac{s_e^2 * \sum_{i=1}^{n} x_i^2}{n * \sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{\frac{0.0016914449298385 * 10.5402}{1.20864 * 10}} = 0.0384065127359383$$
$$se_{\hat{\beta}_2} = \sqrt{\frac{s_e^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \sqrt{\frac{0.0016914449298385}{1.20864}} = 0.0374093747612531$$
$$t_{\beta_1} = \frac{\hat{\beta}_1}{se_{\hat{\beta}_1}} = \frac{0.195037728355838}{0.0384065127359383} = 5.07824622601923$$
$$t_{\beta_2} = \frac{\hat{\beta}_2}{se_{\hat{\beta}_2}} = \frac{0.208035477892507}{0.0374093747612531} = 5.56105198817652$$
$$t_{crit} = t_{0.975,8} = 2.30600413520417$$

$|t_{\beta_2}| > |t_{\beta_1}| > t_{crit}$, следовательно, гипотеза о незначимости коэффициента отвергается в обоих случаях на 5% уровне значимости.

6

$$x_0 = 1.04$$
$$\hat{y}_0 = 0.195037728355838 + 0.208035477892507 * 1.04 = 0.411394625364045$$
$$s_{\hat{y}_0}^2 = s_e^2 * \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right) = 0.0016914449298385^2 * \left(1 + \frac{1}{10} + \frac{(1.04 - 0.966)^2}{1.20864}\right) = 0.0454262333504169$$
$$e_0 = y_0 - \hat{y}_0 = 0.46 - 0.411394625364045 = 0.04860537463$$

95% доверительный интервал прогноза индивидуального значения:

$(\hat{y_0} + t_{0.025,8} * s_{\hat{y_0}}; \hat{y_0} + t_{0.975,8} * s_{\hat{y_0}}) = (0.411394625364045 - 2.30600413520417 * \sqrt{0.04542623};$

$0.411394625364045 + 2.30600413520417 * \sqrt{0.04542623}) = (0.411394625364045 - 0.0996730874667656; 0.411394625364045 + 0.0996730874667656) =$

$= (0.311721512533234; 0.511067687466766)$

Согласно оцененной модели при средней з/п в 1.04 тыс. руб. прожиточный минимум на душу населения равен 411.4 руб.

7

$x_0 = 9$

$\hat{y_0} = 0.195037728355838 + 0.208035477892507 * 9 = 2.06735702939$

$s_M^2 = s_e^2 * \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right) = 0.0016914449298385^2 *$

$\left(\frac{1}{10} + \frac{(9 - 0.966)^2}{1.20864}\right) = 0.0904975937101495$

$s_{\hat{y_0}}^2 = s_e^2 * \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right) = 0.0016914449298385^2 *$

$\left(1 + \frac{1}{10} + \frac{(9 - 0.966)^2}{1.20864}\right) = 0.092189038639988$

95% доверительный интервал:

$(E(y_0|x_0 = 9) + t_{0.025,8} * s_M^2; E(y_0|x_0 = 9) + t_{0.975,8} * s_M^2) = (2.06735702939 - 2.30600413520417 * \sqrt{0.0904975937101495};$

$2.06735702939 + 2.30600413520417 * \sqrt{0.0904975937101495}) =$

$= (1.373646; 2.761068)$

$(\hat{y_0} + t_{0.025,8} * s_{\hat{y_0}}; \hat{y_0} + t_{0.975,8} * s_{\hat{y_0}}) = (2.06735702939 - 2.30600413520417 * \sqrt{0.09218903863998}$

$2.06735702939 + 2.30600413520417 * \sqrt{0.092189038639988}) =$

$= (1.367193; 2.767521)$

Доверительный интервал для индивидуального прогноза шире, чем для условного математического ожидания, так как в первом также учиитывается дополнительно отклонения значений y_0 от своего математического ожидания $E(y_0|x = 9)$.

Задача 2

1

$$X = \begin{bmatrix} 1 & 49 & 16 \\ 1 & 20 & 44 \\ 1 & 31.9 & 13 \\ 1 & 33.4 & 12 \\ 1 & 35.3 & 12 \\ 1 & 24.6 & 18 \\ 1 & 30.8 & 22 \\ 1 & 43.4 & 8 \\ 1 & 42.4 & 10 \\ 1 & 53.8 & 7 \\ 1 & 60.6 & 7 \\ 1 & 58.1 & 6 \end{bmatrix} \tag{1}$$

$$Y = \begin{bmatrix} 71 \\ 67 \\ 72 \\ 71 \\ 72 \\ 73 \\ 73 \\ 78 \\ 72 \\ 77 \\ 76 \\ 77 \end{bmatrix} \tag{2}$$

$$X^T X = \begin{bmatrix} 12 & 483.3 & 175 \\ 483.3 & 21357.79 & 5944.1 \\ 175 & 5944.1 & 3755 \end{bmatrix} \tag{3}$$

$$(X^T X)^{-1} = \begin{bmatrix} 3.53352 & -0.06100303 & -0.06811128 \\ -0.06100303 & 0.001136855 & 0.001043395 \\ -0.06811128 & 0.001043395 & 0.00178893 \end{bmatrix} \tag{4}$$

$$X^T Y = \begin{bmatrix} 879 \\ 35732.9 \\ 12533 \end{bmatrix} \tag{5}$$

$$\hat{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = (X^T X)^{-1} X^T Y = \begin{bmatrix} 72.51012 \\ 0.07834808 \\ -0.1656405 \end{bmatrix} \tag{6}$$

Уравнение регрессии имеет вид:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} = 72.51012 + 0.07834808 * x_{1i} - 0.1656405 x_{2i}$$

Константа в регрессии отражает усредненное влияние всех неучтенных факторов. В среднем при прочих равных при увеличении x_1 на 1 значение y вырастет на 0.07834808, т.е. x_1 положительно влияет на y. В среднем при прочих равных при увеличении x_2 на 1 значение y снижается на 0.1656405, т.е. x_1 отрицательно влияет на y.

2

$$e = y - \beta^T * x = \begin{bmatrix} -2.69893 \\ 0.2110998 \\ -0.856099 \\ -2.139262 \\ -1.288123 \\ 1.544045 \\ 1.720849 \\ 3.414695 \\ -2.175675 \\ 1.434235 \\ -0.09853209 \\ 0.9316976 \end{bmatrix}$$

$$\hat{\sigma}_e = e^T * e * \frac{1}{12 - 1 - 2} = 4.330144$$

$$\widehat{Var}(\hat{\beta}) = \hat{\sigma}_e * (X^T * X)^{-1} = 4.330144 * \begin{bmatrix} 3.53352 & -0.06100303 & -0.06811128 \\ -0.06100303 & 0.001136855 & 0.001043395 \\ -0.06811128 & 0.001043395 & 0.00178893 \end{bmatrix} = \quad (7)$$

$$= \begin{bmatrix} 15.30065 & -0.2641519 & -0.2949316 \\ -0.2641519 & 0.004922747 & 0.004518049 \\ -0.2949316 & 0.004518049 & 0.007746324 \end{bmatrix} \quad (8)$$

$se_{\hat{\beta}_0} = \sqrt{\widehat{Var}(\hat{\beta})_{11}} = \sqrt{15.30065} = 3.911604$
$se_{\hat{\beta}_1} = \sqrt{\widehat{Var}(\hat{\beta})_{22}} = \sqrt{0.004922747} = 0.07016229$
$se_{\hat{\beta}_2} = \sqrt{\widehat{Var}(\hat{\beta})_{33}} = \sqrt{0.007746324} = 0.08801320$
$t_{\beta_1} = \frac{\hat{\beta}_0}{se_{\hat{\beta}_0}} = \frac{72.51012}{3.91160} = 18.537$
$t_{\beta_1} = \frac{\hat{\beta}_1}{se_{\hat{\beta}_1}} = \frac{0.07835}{0.07016} = 1.117$
$t_{\beta_2} = \frac{\hat{\beta}_2}{se_{\hat{\beta}_2}} = \frac{-0.16564}{0.08801} = -1.882$
$t_{crit} = t_{0.975,9} = 2.2621571627982$

$|t_{\beta_0}| > t_{crit}$, следовательно, гипотеза о незначимости константы отвергается на 5% уровне значимости.

$|t_{\beta_1}| < t_{crit}$, следовательно, гипотеза о незначимости коэффициента не отвергается на 5% уровне значимости.

$|t_{\beta_2}| < t_{crit}$, следовательно, гипотеза о незначимости коэффициента не отвергается на 5% уровне значимости.

3

95% доверительные интервал для β_0:

$(\hat{\beta}_0 + t_{0.025,9} * s_{\hat{\beta}_0}; \hat{\beta}_0 + t_{0.975,9} * s_{\hat{\beta}_0} = (72.51012 - 2.2621571627982 * 3.91160; 72.51012 + 2.2621571627982 * 3.91160) = (63.66145876; 81.3587859)$

95% доверительные интервал для β_1:

$(\hat{\beta}_1 + t_{0.025,9} * s_{\hat{\beta}_1}; \hat{\beta}_1 + t_{0.975,9} * s_{\hat{\beta}_1} = (0.07835 - 2.2621571627982 * 0.07016; 0.07835 + 2.2621571627982 * 0.07016) = (-0.08037006; 0.23706621)$

95% доверительные интервал для β_2:

$(\hat{\beta}_2 + t_{0.025,9} * s_{\hat{\beta}_2}; \hat{\beta}_2 + t_{0.975,9} * s_{\hat{\beta}_2} = (-0.16564 - 2.2621571627982 * 0.08801; -0.16564 + 2.2621571627982 * 0.08801) = (-0.36474024; 0.03345916)$

0 принадлежит 95-% доверительному интервалу для β_0 – коэффициент значимы на 5% уровне. 0 принадлежит 95-% доверительному интервалу для β_1, β_2 – эти коэффициенты незначимы на 5% уровне.

4

$$R^2 = 1 - \frac{e^T e}{(Y - \bar{Y})^T (Y - \bar{Y})} = 1 - \frac{38.9712916989129}{112.25} = 0.652817000455118$$

65,3 % дисперсии результативного признака y объясняется влиянием факторных признаков x_1, x_2.

5

$$F_{obs} = \frac{R^2 * (n - 1 - 2)}{(1 - R^2)2} = \frac{0.652817000455118 * 9}{(1 - 0.652817000455118) * 2} = 8.46146414397881$$

$$F_{crit} = F_{2,9,0.05} = 4.25649472909375$$

$F_{obs} > F_{crit}$ коэффициент детерминации является статистически значимым, следовательно, рассматриваемая регрессия достаточно определена включенными переменными.

6

Проведем тест ранговой корреляции Спирмена для переменной x_1:

$$d_1 = r_e - r_{x1} = \begin{bmatrix} 1 \\ 7 \\ 5 \\ 3 \\ 4 \\ 10 \\ 11 \\ 12 \\ 2 \\ 9 \\ 6 \\ 8 \end{bmatrix} - \begin{bmatrix} 9 \\ 1 \\ 4 \\ 5 \\ 6 \\ 2 \\ 3 \\ 8 \\ 7 \\ 10 \\ 12 \\ 11 \end{bmatrix} = \begin{bmatrix} -8 \\ 6 \\ 1 \\ -2 \\ -2 \\ 8 \\ 8 \\ 4 \\ -5 \\ -1 \\ -6 \\ -3 \end{bmatrix} \tag{9}$$

$\rho_1 = 1 - \dfrac{6 * d_1^T * d}{n(n^2 - n)} = 1 - \dfrac{6 * 324}{12(12^2 - 12)} = -0.132867132867133$
$t_{obs} = \dfrac{\rho_1 \sqrt{n-2}}{\sqrt{1-\rho_1^2}} = \dfrac{-0.132867132867133\sqrt{12-2}}{\sqrt{1 - 0.132867132867133^2}} = -0.423921312476319$
$t_{crit} = t_{0.975,10} = 2.22813885198627$

$|t_{obs}| < t_{crit}$, следовательно, нулевая гипотеза об отстутствии корреляции между вектором ошибок и x_1 не отвергается.

Проведем тест ранговой корреляции Спирмена для переменной x_2:

$$d_2 = r_e - r_{x2} = \begin{bmatrix} 1 \\ 7 \\ 5 \\ 3 \\ 4 \\ 10 \\ 11 \\ 12 \\ 2 \\ 9 \\ 6 \\ 8 \end{bmatrix} - \begin{bmatrix} 9 \\ 12 \\ 8 \\ 6.5 \\ 6.5 \\ 10 \\ 11 \\ 4 \\ 5 \\ 2.5 \\ 2.5 \\ 1 \end{bmatrix} = \begin{bmatrix} -8 \\ -5 \\ -3 \\ -3.5 \\ -2.5 \\ 0 \\ 0 \\ 8 \\ -3 \\ 6.5 \\ 3.5 \\ 7 \end{bmatrix} \tag{10}$$

$\rho = 1 - \dfrac{6 * d_2^T * d}{n(n^2 - n)} = 1 - \dfrac{6 * 293}{12(12^2 - 12)} = -0.0244755244755244$
$t_{obs} = \dfrac{\rho_1 \sqrt{n-2}}{\sqrt{1 - \rho_1^2}} = \dfrac{-0.0244755244755244\sqrt{12-2}}{\sqrt{1 - 0.0244755244755244^2}} = -0.0774215974981098$
$t_{crit} = t_{0.975,10} = 2.22813885198627$

$|t_{obs}| < t_{crit}$, следовательно, нулевая гипотеза об отстутствии корреляции между вектором ошибок и x_2 не отвергается.

Вывод – ошибки гомоскедастичны.

7

$DW = \dfrac{\sum_2^{12}(e_t - e_{t-1})^2}{\sum_1^{12} e_t^2} = ((0.211099759340403 + 2.6989295510695)^2 + (-0.85609903244597$

$0.211099759340403)^2 + (-2.13926168682379 + 0.856099032445978)^2 + (-1.28812303475612 +$

$2.13926168682379)^2 + (1.54404462387122 + 1.28812303475612)^2 + (1.72084869150117 - 1.5440446238$

$(3.41469538422851 - 1.72084869150117)^2 + (-2.17567546273412 - 3.41469538422851)^2 +$

$(1.43423483690452 + 2.17567546273412)^2 + (-0.0985320925375037 - 1.43423483690452)^2 +$

$(0.931697564521175 + 0.0985320925375037)^2)/((-2.6989295510695)^2 + (0.211099759340403)^2 +$

$(-0.856099032445978)^2 + (-2.13926168682379)^2 + (-1.28812303475612)^2 + (1.54404462387122)^2 +$

$(1.72084869150117)^2 + (3.41469538422851)^2 + (-2.17567546273412)^2 + (1.43423483690452)^2 +$

$(-0.0985320925375037)^2 + (0.931697564521175)^2) = \dfrac{70.5941233789183}{38.9712916989129} = 1.81143914664977$

$k = 2, n = 12$

$d_{U,0.05} = 1.274$

$d_{L,0.05} = 0.569$

$DW > d_U$, следовательно, гипотеза об отсутствии автокорреляции не отвергается.