Инжењерска статистика
Вежба 9

Једноставна линеарна регресија и корелација

- **Регресиона и корелациона анализа је статистичка метода која се користи за моделирање и испитивање односа између две или више променљивих.**

- **У оквиру овог поглавља приказана су испитивања која се односе на једностуку линеарну регресиону и корелациону анализу.**

- **То значи да се посматра однос две или више променљиве - *X и Y* и испитује се да ли је њихова међусобна зависност линеарна**

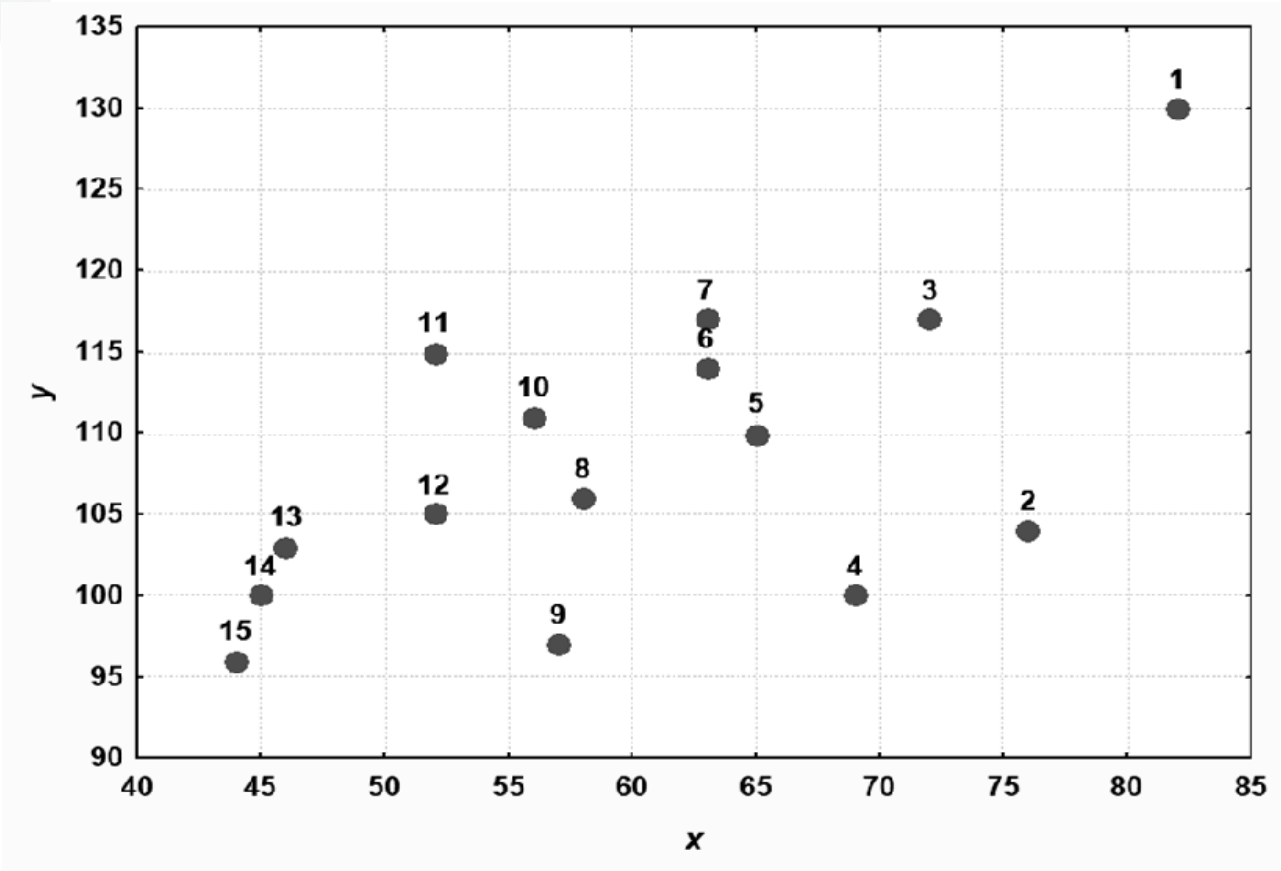
Једноставна линеарна регресија и корелација

- **Испитивање се врши преко парова мерења**

$$(x_1, y_1), \cdots, (x_n, y_n)$$

- **Први корак је конструисање *дијаграма расипања,* односно дводимензионалног координатног система у који се уноси измерених *n* парова тачака.**

Дијаграм расипања



- На основу дијаграма расипања претпоставља се да је регресиона зависност линеарна или се може прилагодити линеарној, односно да је облика:

$$y_i = a \cdot x_i + b + e_i$$

- Где су:
  - x_i и y_i променљиве чија се међусобна веза испитује, где је x_i независна променљива, а y_i зависна
  - *a* и *b* параметри регресионе праве или *регресиони коефицијенти*
  - e_i случајна грешка или *резидуал (residual)*
- Претпоставља се да се резидуали распоређују независно, по нормалној расподели са средином 0 и варијансом σ^2
- Да би се поставила регресиона права потребно је *оценити параметре a и b*. Ова оцена се врши преко методе *најмањих квадрата.*

- **Преко методе најмањих квадрата добија се систем једначина за овену параметара линеарне регресионе зависности**

$$\sum_{i=1}^{n} y_i = a\sum_{i=1}^{n} x_i + n\cdot b$$

$$\sum_{i=1}^{n} y_i \cdot x_i = a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i$$

$$: n \Rightarrow \begin{aligned} \bar{y} &= a\bar{x} + b \\ \overline{xy} &= a\overline{x^2} + b\bar{x} \end{aligned}$$

$$\hat{a} = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\overline{x^2} - (\bar{x})^2} \qquad \hat{b} = \bar{y} - \hat{a}\bar{x}$$

(оцене параметара регресионе праве)

Где су:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

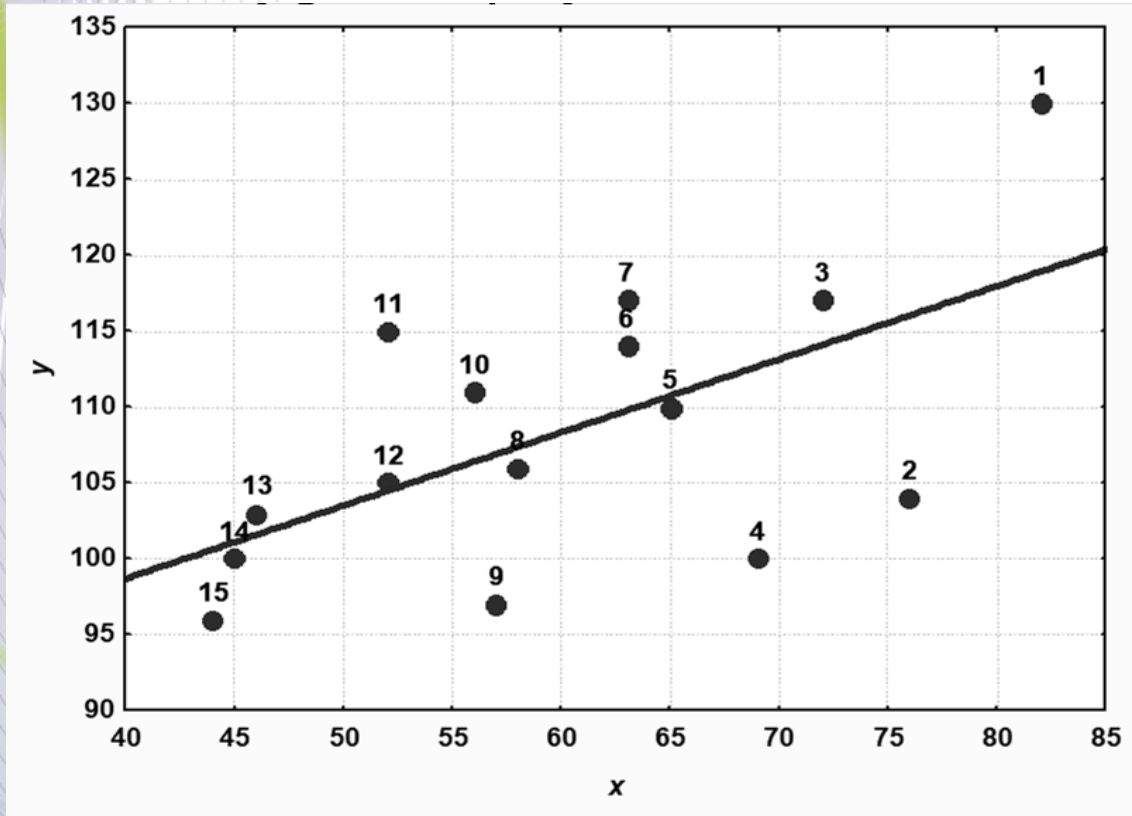
$$\overline{y^2} = \frac{1}{n}\sum_{i=1}^{n} y_i^2$$

$$\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} y_i x_i$$

- **Оцењене вредности регресионе линије**

$$\hat{y} = \hat{a}x + \hat{b}$$

и уноси се у дијаграм расипања



Постављену регресиону зависност је потребно потврдити. То се ради на два начина, зависно од тога какве су променљиве:

- **Преко коефицијента корелације *r*, у случају да су обе променљиве случајне**
- **Преко анализе варијансе, ако је једна променљива детерминистичка**

Коефицијент корелације

- ***Коефицијент корелације*** **показује јачину везе тачака које се расипају око регресионе праве.**
- **Вредност му се креће *искључиво* у границама (-1,+1).**
- **Рачунска вредност се добија:**

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\left[\overline{x^2} - (\bar{x})^2\right] \cdot \left[\overline{y^2} - (\bar{y})^2\right]}}$$

- **Јачина корелационе зависности се смањује приближавањем коефицијената корелације нули.**

У пракси је често потребно донети одлуку на основу коефицијената корелације. Практичне препоруке за одређивање постојања регресионе зависности преко коефицијената корелације су:

$|r| \in [0.9; 1]$ **Веза је апсолутна,**

$|r| \in [0.7; 0.9)$ **Веза је врло јака,**

$|r| \in [0.5; 0.7)$ **Веза је слаба и**

$|r| \in [0; 0.5)$ **Не постоји корелациона веза**

Тестирање постојања регресионе зависности преко анализе варијансе

Утицајни фактор који се испитује је регресиона сума квадрата, а табела за анализу варијансе је:

izvor varijacije	*SK*	*f*	ocena varijanse	F_0
regresija	$SK_r = \hat{a}\left(\overline{xy} - \bar{x}\cdot\bar{y}\right)$	1	$S_r^2 = \frac{SK_r}{f_r}$	$\frac{S_r^2}{S_e^2}$
greška	$SK_e = SK_T - SK_r$	*n*-2	$S_e^2 = \frac{SK_e}{f_e}$	
total	$SK_T = \overline{y^2} - \left(\bar{y}\right)^2$	*n*-1	$F_0 > F_{\alpha,1,n-2}$	

Без обзира да ли се за доказивање регресионе зависности користи коефицијент корелације или анализа варијансе, тумачење регресионе зависности се врши преко *коефицијента детерминације.*

Коефицијент детерминације - $r^2(\%)$ представља проценат од укупне варијације који се може објаснити постављеном регресионом зависношћу

Предвиђање нове (до сад не измерене) вредности зависне променљиве у односу на познату вредност зависне се врши помоћу формуле:

$$\hat{y}_j = \hat{a}\cdot x_j + \hat{b}$$

Процедура решавања

1. **Дефинисање промељивих**
2. **Конструисање дијаграма расипања**
3. **Претпостављање линеарне зависности**
4. **Одређивање параметара регресионе праве**
5. **Потврђивање регресионе зависности – преко *r* или ANOVA**
6. **Одређивање и тумачење коефицијента детерминације**
7. **Предвиђање *(алтернативно/опционо)***

Пример 1

Петнаест студената група за Индустријско инжењерство изабрано је на случајан начин. Код студената су поређени бодови које су добили на првим колоквијумима из Инжењерске статистике и Операционих истраживања и резултати су приказани у табели. Потребно је: Утврдити да ли постији зависност између података, потврдити ту зависност, извршити додатна испитивања која су потребна и предвидети који број бодова из ОИ се може очекивати од студента који је на ИС добио 7.2.

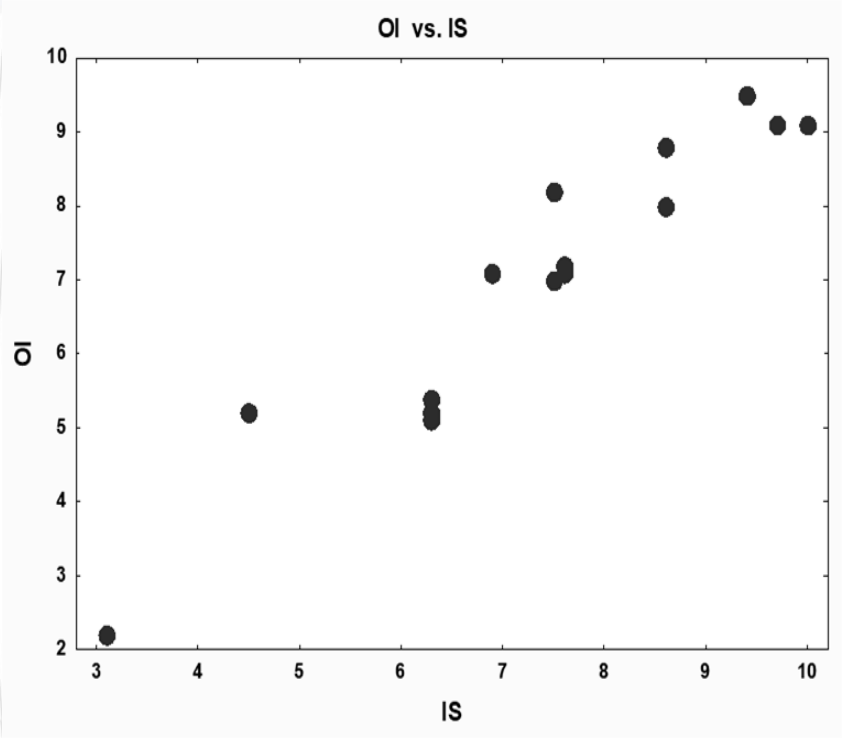
ИС	ОИ	ИС	ОИ	ИС	ОИ
8.6	8	7.5	8.2	6.9	7.1
10	9.1	6.3	5.2	9.4	9.5
8.6	8.8	7.6	7.1	7.6	7.2
4.5	5.2	6.3	5.4	3.1	2.2
9.7	9.1	7.5	7	6.3	5.1

1. Дефинисање променљивих

x – оцена из инжењерске статистике ИС - случајна, независна

y – оцена из операционих истраживања ОИ – случајна, зависна

Конструисање дијаграма расипања



3. Претпостављање линеарне зависност

$$y_i = a \cdot x_i + b + e_i$$

4. Одређивање параметара регресионе праве

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{15}(8.6 + \cdots + 6.3)$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{15}(8 + \cdots + 5.1)$$

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{1}{15}(8.6^2 + \cdots + 6.3^2)$$

$$\overline{y^2} = \frac{1}{n}\sum_{i=1}^{n} y_i^2 = \frac{1}{15}(8^2 + \cdots + 5.1^2)$$

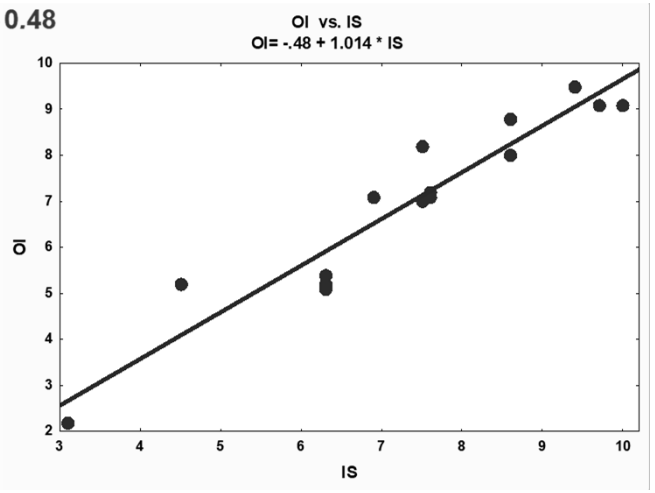
$$\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} y_i x_i = \frac{1}{15}(8.6 \cdot 8 + \cdots + 6.3 \cdot 5.1)$$

Мерење	x	y	x^2	y^2	xy
1	8.6	8	73.96	64	68.8
2	10	9.1	100	82.81	91
3	8.6	8.8	73.96	77.44	75.68
4	4.5	5.2	20.25	27.04	23.4
5	9.7	9.1	94.09	82.81	88.27
6	7.5	8.2	56.25	67.24	61.5
7	6.3	5.2	39.69	27.04	32.76
8	7.6	7.1	57.76	50.41	53.96
9	6.3	5.4	39.69	29.16	34.02
10	7.5	7	56.25	49	52.5
11	6.9	7.1	47.61	50.41	48.99
12	9.4	9.5	88.36	90.25	89.3
13	7.6	7.2	57.76	51.84	54.72
14	3.1	2.2	9.61	4.84	6.82
15	6.3	5.1	39.69	26.01	32.13
Σ	109.9	104.2	854.93	780.3	813.85
Средина	7.327	6.947	56.995	52.02	54.257

$$\hat{a} = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\overline{x^2} - (\bar{x})^2} = \frac{54.257 - 7.327\cdot 6.947}{56.995 - 7.327^2} = \frac{3.361}{3.153} = 1.014$$

$$\hat{b} = \bar{y} - \hat{a}\bar{x} = 6.647 - 1.014\cdot 7.327 = -0.480$$

$$\hat{y}_i = 1.014x_i - 0.48$$



5. **Потврђивање регресионе зависности**

$$r = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\sqrt{\left[\overline{x^2} - (\bar{x})^2\right]\cdot\left[\overline{y^2} - (\bar{y})^2\right]}} \qquad r = \frac{54.257 - 7.327\cdot 6.947}{\sqrt{\left(56.995 - 7.327^2\right)\left(52.02 - 6.947^2\right)}} = 0.951$$

$r = 0.9514 \wedge |r| \in [0.9;1] \Rightarrow$ **Корелациона веза је апсолутна**

6. **Коефицијент детерминације**

$$r^2 = 0.951^2\cdot 100(\%) = 90.52\%$$

7. **Предвиђање**

Када је оцена из ИС 7.2, предвиђена оцена из ОИ је:

$$\hat{y}_{7.2} = 1.014\cdot 7.2 - 0.48 = 6.8$$

Пример 2

Радник је добио да производи нови производ. Током 6 дана мерен је број добрих производа које је направио. Подаци су приказани у табели. Поставити одговарајући регресиони модел и одредити коефицијент корелације.

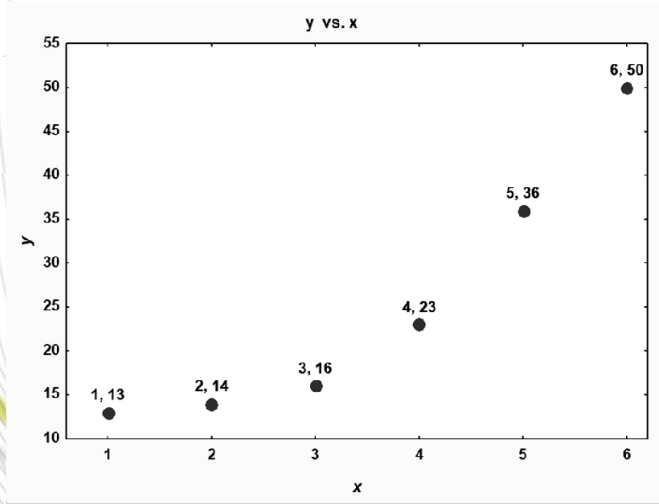
дан	1	2	3	4	5	6
Добри производи	13	14	16	23	36	50

1. Дефинисање променљивих

x **- дан – детерминистичка - независна**

y **– број добрих производа- случајна - зависна**

2. Конструисање дијаграма расипања



3. Претпостављање линеарне зависности

$$y_i = a \cdot x_i + b + e_i$$

Трансформација

$$y = b \cdot e^{a \cdot x} \qquad Y = ln(y) \qquad B = ln(b) \qquad \Rightarrow Y_i = B + a \cdot x_i + e_i$$

$$y = e^{ax+b}$$

$$y = ln(ax+b)$$

$$e^y = ax + b$$

$$y = ax^b$$

4. Одређивање параметара регресионе праве

	x	y	Y	x^2	Y^2	xY
1	1	13	2.565	1	6.579	2.565
2	2	14	2.639	4	6.965	5.278
3	3	16	2.772	9	7.687	8.318
4	4	23	3.135	16	9.831	12.542
5	5	36	3.584	25	12.842	17.918
6	6	50	3.912	36	15.304	23.472
Σ	21		18.608	91	59.208	70.092
средина	3.5		3.101	15.167	9.868	11.682

$$\hat{a} = \frac{\overline{xY} - \bar{x}\bar{Y}}{\overline{x^2} - (\bar{x})^2} = \frac{11.68 \quad -3.5 \cdot 3.101}{15.167 - 3.5^2} = 0.284$$

$$\hat{B} = \bar{Y} - \hat{a}\bar{x} = 3.101 - 0.284 \cdot 3.5 = 2.108$$

$$\hat{b} = e^{\hat{B}} = e^{2.108} = 8.232$$

$$\hat{y}_i = 8.232 \cdot e^{0.284 \cdot x_i}$$

5. Потврђивање регресионе зависности

izvor varijacije	*SK*	*f*	ocena varijanse	F_0
regresija	1.409	1	1.409	61.766**
greška	0.091	4	0.0234	
total	1.500	5		

6. Коефицијент детерминације

$$r^2 = 0.932^2 \cdot 100(\%) = 86.8\%$$

Вишеструка линеарна регресија и корелација

- Вишеструка регресиона и корелациона анализа је статистичка метода која се користи за моделирање и испитивање односа више променљивих.

- То значи да се посматра међусобни однос и зависност више променљивих Y – *зависне* и $X_1,\ldots, X_m$ - *независне* m=1,2,.... Претпоставља се да је зависност линеарна.

- Испитивање се врши преко n парова опсервација, где се зависна променљива пореди са свим независним. Могуће је радити поређење зависне променљиве и појединих независних.

- Код вишеструке линеарне регресије само се у појединим случајевима црта *дијаграм расипања*

- **Претпоставља се да је регресиона зависност линеарна или се може прилагодити линеарној, облика**

$$y_i = a \cdot x_{1i} + \ldots + g \cdot x_{ji} + \ldots + r \cdot x_{mi} + s + e_i$$

Где су:

- ***x_{ji}* и *y_i* променљиве чија се међусобна веза испитује, где су *x_{ji}* независне промељиве, док је *y_i* зависнаа** $i = \overline{1,n} \quad j = \overline{1,m}$
- ***a, b*,c ..., s су параметри регресионе зависности или *регресиони коефицијенти***
- ***e_i* случајна грешка или *резидуал***

- **Претпоставља се да се резидуали распоређују независно, по нормалној расподели са средином 0 и варијансом** σ^2

- **Да би се поставила регресиона зависност потребно је *оценити регресионе коефицијенте, методом најмањих квадрата.***

$$\sum_{i=1}^{n} y_i = a\sum_{i=1}^{n} x_{1i} + \ldots + g\sum_{i=1}^{n} x_{ji} + \ldots + r\sum_{i=1}^{n} x_{mi} + n\cdot s \quad (1)$$

$$\sum_{i=1}^{n} y_i = a\sum_{i=1}^{n} x_{1i} + \ldots + g\sum_{i=1}^{n} x_{ji} + \ldots + r\sum_{i=1}^{n} x_{mi} + n\cdot s \quad / \cdot x_{1i}$$

$$\sum_{i=1}^{n} y_i x_{1i} = a\sum_{i=1}^{n} (x_{1i})^2 + \ldots + r\sum_{i=1}^{n} x_{mi}\cdot x_{1i} + s\sum_{i=1}^{n} x_{1i} \quad (2)$$

$$\bullet\ \bullet\ \bullet$$

$$\sum_{i=1}^{n} y_i x_{mi} = a\sum_{i=1}^{n} x_{1i}\cdot x_{mi} + \ldots + r\sum_{i=1}^{n} (x_{mi})^2 + s\sum_{i=1}^{n} x_{mi} \quad (m)$$

$$\sum_{i=1}^{n} y_i = a\sum_{i=1}^{n} x_{1i} + \ldots + r\sum_{i=1}^{n} x_{mi} + n\cdot s$$

$$\sum_{i=1}^{n} y_i x_{1i} = a\sum_{i=1}^{n} (x_{1i})^2 + \ldots + r\sum_{i=1}^{n} x_{mi}\cdot x_{1i} + s\sum_{i=1}^{n} x_{1i}$$

$$\bullet\ \bullet\ \bullet$$

$$\sum_{i=1}^{n} y_i x_{mi} = a\sum_{i=1}^{n} x_{1i}\cdot x_{mi} + \ldots + r\sum_{i=1}^{n} (x_{mi})^2 + s\sum_{i=1}^{n} x_{mi} \quad / : n$$

$$\overline{y} = a\cdot\overline{x_1} + \ldots + g\cdot\overline{x_j} + \ldots + r\cdot\overline{x_m} + s$$

$$\overline{yx_1} = a\cdot\overline{x_1^2} + \ldots + r\cdot\overline{x_m x_1} + s\cdot\overline{x_1}$$

$$\bullet\ \bullet\ \bullet$$

$$\overline{yx_m} = a\cdot\overline{x_1 x_m} + \ldots + r\cdot\overline{x_m^2} + s\cdot\overline{x_m}$$

Где су:

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \overline{x_1} = \frac{1}{n}\sum_{i=1}^{n} x_{1i} \qquad \overline{x_m} = \frac{1}{n}\sum_{i=1}^{n} x_{mi}$$

$$\overline{y^2} = \frac{1}{n}\sum_{i=1}^{n} y_i^2 \qquad \overline{x_1^2} = \frac{1}{n}\sum_{i=1}^{n} x_{1i}^2 \qquad \bullet\ \bullet\ \bullet \qquad \overline{x_m^2} = \frac{1}{n}\sum_{i=1}^{n} x_{mi}^2$$

$$\overline{yx_1} = \frac{1}{n}\sum_{i=1}^{n} y_i x_{1i} \qquad \overline{yx_m} = \frac{1}{n}\sum_{i=1}^{n} y_i x_{mi}$$

$$\bullet\ \bullet\ \bullet \qquad \bullet\ \bullet\ \bullet$$

$$\overline{x_1 x_m} = \frac{1}{n}\sum_{i=1}^{n} x_{1i} x_{1m} \qquad \overline{x_1 x_m} = \frac{1}{n}\sum_{i=1}^{n} x_{1i} x_{1m}$$

Постављену регресиону зависност је потребно и потврдити. То се ради на два начина, зависно од тога какве су променљиве:

- ☒ **Преко коефицијента корелације *r*, у случају да су све променљиве случајне**
- ☒ **Преко анализе варијансе, ако је једна променљива детерминистичка**

Коефицијент корелације **показује јачину везе између свих параметара регресионе зависности. Могу се рачунати и поједини парцијални коефицијенти корелације.**

Вредност му се креће *искључиво* у границама (-1,+1).

Рачунска вредност:

$$r = \sqrt{\frac{SK_r}{SK_T}}$$

Јачина корелационе зависности смањује се приближавањем коефицијента корелације нули.

У пракси је често потребно донети одлуку на основу коефицијента корелације.

Практичне препоруке за одређивање регресионе зависности преко коефицијената корелације су:

$|r| \in [0.9;1]$ **Веза је апсолутна,**

$|r| \in [0.7;0.9)$ **Веза је јака,**

$|r| \in [0.5;0.7)$ **Веза је слаба и**

$|r| \in [0;0.5)$ **Не постоји корелациона веза**

Тестирање постављене корелационе зависности преко анализе варијансе

Утицајни фактор који се испитује је регресиона сума квадрата, а табела анализе варијансе је:

Извор вар.	*SK*	*f*	Оцена вар.	F_0
Регресија	$SK_r = SK_T - SK_e$	*m*	$S_r^2 = \frac{SK_r}{f_r}$	$\frac{S_r^2}{S_e^2}$
Грешка	$SK_e = \sum_{i=1}^{n} e_i^2$	*n-m-1*	$S_e^2 = \frac{SK_e}{f_e}$	
Тотал	$SK_T = \overline{y^2} - (\bar{y})^2$	*n*-1	$e_i = y_i - \hat{y}_i$	

Регресиона зависност постоји уколико је: $F_0 > F_{\alpha,m,n-m-1}$

Без обзира да ли се за доказивање регресионе зависности користи коефицијент корелације или анализа варијансе, тумачење регресионе зависности се врши преко *коефицијента детерминације*

Коефицијент детерминације - r^2(%) **представља проценат од укупне варијације који се може објаснити постављеном регресионом зависношћу**

Предвиђање нове зависне случајне променљиве

$$\hat{y}_j = \hat{a} \cdot x_{1j} + \ldots + \hat{r} \cdot x_{mj} + \hat{s}$$

Предвиђање

1. **Дефинисање променљивих**
2. **Предпостављање линеарне зависности**
3. **Одређивање параметара регресионе праве**
4. **Потврђивање регресионе зависности преко *r* или ANOVA**
5. **Одређивање и тумачење коефицијента детерминације**
6. **Предвиђање (опционо)**

Пример 3

Спровести регресиону анализу и одредити коефицијент корелације за податке приказане у табели:

y	10	11	13	14	15	16	17	19
x_1	1	1	3	3	7	7	9	9
x_2	5	5	7	7	9	9	8	8
x_3	3	6	2	1	7	8	5	4

1. Дефинисање променљивих

x_1, x_2 и x_3 – случајне, независне променљиве

y – случајна, зависна променљива

Предпоставка вишеструке регресионе зависности облика:

$$y_i = ax_{1i} + bx_{2i} + cx_{3i} + d + e_i$$

3. Оцена регресионих коефицијената

поставка система једначина

$$\sum_{i=1}^{8} y_i = a\sum_{i=1}^{8} x_{1i} + b\sum_{i=1}^{8} x_{2i} + c\sum_{i=1}^{8} x_{ni} + dn$$

$$\sum_{i=1}^{8} y_i x_{1i} = a\sum_{i=1}^{8} x_{1i}^2 + b\sum_{i=1}^{8} x_{1i}x_{2i} + c\sum_{i=1}^{8} x_{1i}x_{3i} + d\sum_{i=1}^{8} x_{1i}$$

$$\sum_{i=1}^{8} y_i x_{2i} = a\sum_{i=1}^{8} x_{1i}x_{2i} + b\sum_{i=1}^{8} x_{2i}^2 + c\sum_{i=1}^{8} x_{2i}x_{3i} + d\sum_{i=1}^{n} x_{2i}$$

$$\sum_{i=1}^{8} y_i x_{3i} = a\sum_{i=1}^{8} x_{1i}x_{3i} + b\sum_{i=1}^{8} x_{2i}x_{3i} + c\sum_{i=1}^{8} x_{3i}^2 + d\sum_{i=1}^{n} x_{3i}$$

$$/\,n$$

$$\bar{y} = a\bar{x}_1 + b\bar{x}_2 + c\bar{x}_3 + d$$

$$\overline{yx_1} = a\overline{x_1^2} + b\overline{x_1x_2} + c\overline{x_1x_3} + d\bar{x}_1$$

$$\overline{yx_2} = a\overline{x_1x_2} + b\overline{x_2^2} + c\overline{x_2x_3} + d\bar{x}_2$$

$$\overline{yx_3} = a\overline{x_1x_3} + b\overline{x_2x_3} + c\overline{x_3^2} + d\bar{x}_3$$

мер.	y	x_1	x_2	x_3	x_1^2	x_2^2	x_3^2	x_1x_2	x_1x_3	x_2x_3	yx_1	yx_2	yx_3
1	10	1	5	3	1	25	9	5	3	15	10	50	30
2	11	1	5	6	1	25	36	5	6	30	11	55	66
3	13	3	7	2	9	49	4	21	6	14	39	91	26
4	14	3	7	1	9	49	1	21	3	7	42	98	14
5	15	7	9	7	49	81	49	63	49	63	105	135	105
6	16	7	9	8	49	81	64	63	56	72	112	144	128
7	17	9	8	5	81	64	25	72	45	40	153	136	85
8	19	9	8	4	81	64	16	72	36	32	171	152	76
Σ	115	40	58	36	280	438	204	322	204	273	643	861	530
сред	14.4	5	7.25	4.5	35	54.8	25.5	40.2	25.5	34.1	80.4	107.6	66.2

$$5.00a + 7.250b + 4.500c + 1.00d = 14.375$$
$$35.00a + 40.250b + 25.500c + 5.00d = 80.375$$
$$40.25a + 54.750b + 34.125c + 7.25d = 107.625$$
$$25.50a + 34.125b + 25.500c + 4.50d = 66.250$$

$$\hat{a} = 0.868$$
$$\hat{b} = 0.131$$
$$\hat{c} = -0.236$$
$$\hat{d} = 10.144$$

Тако да је регресиона зависност:

$$\hat{y} = 0.868x_1 + 0.131x_2 - 0.236x_3 + 10.144$$

4. Потврђивање регресионе зависности - преко ANOVA или *r*

Укупни коефицијент корелације је:

$$r = \sqrt{\frac{59.671}{63.875}} = 0.966$$

5. Коефицијент детерминације

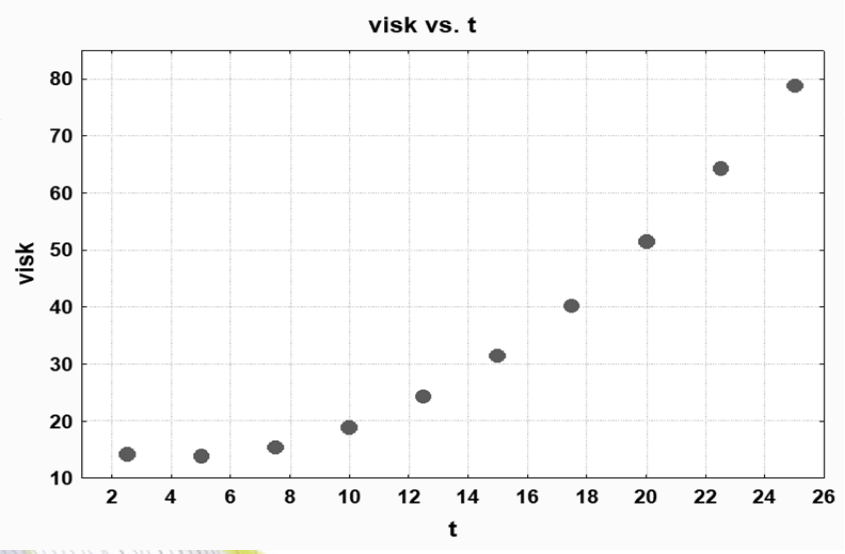
$$r^2 = 93.419\%$$

Пример 4 (једнострука)

Извршено је испитивање промене вискозности уља зависно од температуре. Резултати су приказани у табели. Потребно је према одговарајућој процедури спровести одговарајућу регресиону и корелациону анализу.

Промена вискозности	*t*	Промена вискозности	*t*
14.2	2.5	31.5	15.0
13.9	5.0	40.3	17.5
15.5	7.5	51.5	20.0
18.9	10.0	64.3	22.5
24.3	12.5	78.9	25.0

1. Дефинисање променљивих:
Зависна променљива *y*
Независна променљива *x*



2. Претпоставља се полинимијална зависност са полиномом другог реда:

$$y_i = ax_i + bx_i^2 + c + e_i$$

3. Оцена регресионих коефицијената:

$$\sum_{i=1}^{n} y_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 + nc$$

$$\sum_{i=1}^{n} y_i x_i = a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} y_i x_i^2 = a\sum_{i=1}^{n} x_i^3 + b\sum_{i=1}^{n} x_i^4 + c\sum_{i=1}^{n} x_i^2$$

$$\bar{y} = a\bar{x} + b\overline{x^2} + c$$

$$\overline{yx} = a\overline{x^2} + b\overline{x^3} + c\bar{x}$$

$$\overline{yx^2} = a\overline{x^3} + b\overline{x^4} + c\overline{x^2}$$

Мер.	*y*	*x*	x^2	x^3	x^4	*xy*	x^2y
1	14.2	2.5	6.2	15.6	39.0625	35.5	88.8
2	13.9	5	25	125	625	69.5	347.5
3	15.5	7.5	56.25	421.9	3164.1	116.2	871.9
4	18.9	10	100	1000	10000	189	1890
5	24.3	12.5	156.2	1953.1	24414.06	303.8	3796.9
6	31.5	15	225	3375	50625	472.5	7087.5
7	40.3	17.5	306.2	5359.4	93789.06	705.2	12341.9
8	51.5	20	400	8000	160000	1030	20600
9	64.3	22.5	506.2	11390.6	256289.1	1446.8	32551.9
10	78.9	25	625	15625	390625	1972.5	49312.5
Сума	353.3	137.5	2406.2	47265.6	989570.3	6341	128888.8
Сред.	35.3	13.75	240.6	4726.6	98957	634.1	12888.9

$$13.75a + 240.625b + 1.0c = 35.33$$

$$240.625a + 4726.562b + 13.35c = 634.1$$

$$4726.562a + 98957.0312b + 240.625c = 12888.875$$

$$\hat{a} = -1.244$$

$$\hat{b} = 0.150$$

$$\hat{c} = 16.380$$

$$\hat{y} = -1.244x + 0.15x^2 + 16.38$$

4. Потврђивање регресионе зависности - преко ANOVA

Извор	*SK*	*f*	Оцена варијанс.	F_0
регресија	4728.945	2	2364.472	262719*
грешка	0.056	6	0.009	
тотал	4729.001	9		

$$F_{0.05,2,6} = 5.143$$

5. Коефицијент детерминације

$$r^2 = 99.99\%$$

Домаћи задатак 7

- Решити проблем једноструке линеарне регресије и корелације