

1. Z šesti států máme data o roční spotřebě cigaret na 1 obyvatele (znak x) a o roční míře úmrtnosti na plicní rakovinu na 100 000 obyvatel (znak y). Data jsou zaokrouhlena na stovky, resp. jednotky. Vypočítejte koeficient korelace mezi oběma znaky.

x	3400	2600	2200	2400	2900	2100
y	24	20	17	19	26	20

výpočty pro znak x

aritmetický průměr $\bar{x} = \frac{3400 + 2600 + \dots + 2100}{6} = 2600$

směrodatná odchylka $s_x = \sqrt{\frac{1}{6}[(3400 - 2600)^2 + (2600 - 2600)^2 + \dots + (2100 - 2600)^2]} = 443,471$

výpočty pro znak y

aritmetický průměr $\bar{y} = \frac{24 + 20 + 17 + 19 + 26 + 20}{6} = 21$

směrodatná odchylka $s_y = \sqrt{\frac{1}{6}[(24 - 21)^2 + (20 - 21)^2 + \dots + (20 - 21)^2]} = 3,055$

výpočty pro znaky x i y

aritmetický průměr $\overline{xy} = \frac{3400 \cdot 24 + 2600 \cdot 20 + \dots + 2100 \cdot 20}{6} = \frac{334000}{6} = 55666,667$

korelační koeficient $r_{xy} = \frac{55666,667 - 2600 \cdot 21}{443,471 \cdot 3,055} = \frac{1066,667}{1354,804} = 0,787$

Závislost průměrné spotřeby cigaret a hodnoty počet úmrtí na rakovinu je velká a přímoúměrná. To znamená, čím vyšší je spotřeba cigaret, tím vyšší je počet úmrtí na rakovinu.

y	1	2	3	4	5
x			1	2	1
	2		1	2	3
	4	1	2	4	1
	6		1		

2. Z tabulky rozdělení četností vypočítejte koeficient korelace r_{xy} .

výpočty pro znak x

x_i	0	2	4	6	Σ
n_i	4	7	8	1	20
$x_i \cdot n_i$	0	14	32	6	52
$(x_i - \bar{x})^2$	6,76	0,36	1,96	11,56	---
$(x_i - \bar{x})^2 \cdot n_i$	27,04	2,52	15,68	11,56	56,8

výpočty pro znak y

y_i	1	2	3	4	5	Σ
n_i	1	4	7	6	2	20
$y_i \cdot n_i$	1	8	21	24	10	64
$(y_i - \bar{y})^2$	4,84	1,44	0,04	0,64	3,24	---
$(y_i - \bar{y})^2 \cdot n_i$	4,84	5,76	0,28	3,84	6,48	21,2

výpočty pro znaky x i y

$\Sigma x_i \cdot y_i = 0 \cdot (1 \cdot 3 + 2 \cdot 4 + 1 \cdot 5) + 2 \cdot (1 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 + 1 \cdot 5) + 4 \cdot (1 \cdot 1 + 2 \cdot 2 + 4 \cdot 3 + 1 \cdot 4) + 6 \cdot (1 \cdot 2) = 0 + 50 + 84 + 12 = 146$

aritmetický průměr $\bar{x} = \frac{52}{20} = 2,6$

směrodatná odchylka

$s_x = \sqrt{\frac{1}{20} \cdot 56,8} = \sqrt{2,84} = 1,6852$

aritmetický průměr $\bar{y} = \frac{64}{20} = 3,2$

směrodatná odchylka

$s_y = \sqrt{\frac{1}{20} \cdot 21,2} = \sqrt{1,06} = 1,02956$

$\frac{1}{n} \sum x_i y_i = \frac{1}{20} \cdot 146 = 7,3$

korelační koeficient $r_{xy} = \frac{7,3 - 2,6 \cdot 3,2}{1,6852 \cdot 1,02956} = \frac{-1,02}{1,735015} = -0,588$

Závislost znaků x a y není velká a je nepřímouměrná.

3. 18 žáků 7. třídy ZŠ se podrobilo inteligenčnímu testu. Výsledkem testu je kvocient IQ udávaný v bodech; u převážné většiny populace leží mezi 80 a 120. Výsledky testu byly porovnány s průměrem známek na výročním vysvědčení žáků (žáci jsou již seřazeni podle prospěchu):

známka	IQ	známka	IQ	známka	IQ
1,00	134	1,82	102	2,82	94
1,18	108	2,00	118	2,91	86
1,27	140	2,18	96	3,09	100
1,36	116	2,36	100	3,18	90
1,45	120	2,55	96	3,27	68
1,64	108	2,64	86	3,45	80

Vypočítejte koeficient korelace mezi průměrnou známkou a inteligenčním kvocientem žáků.

výpočty pro znak x

$$\text{aritmetický průměr } \bar{x} = \frac{1,00 + 1,18 + \dots + 3,45}{18} = \frac{40,17}{18} = 2,23$$

$$\text{směrodatná odchylka } s_x = \sqrt{\frac{1}{18} [(2,23 - 1,00)^2 + \dots + (3,45 - 2,23)^2]} = \sqrt{\frac{1}{18} \cdot 10,7839} \doteq 0,774$$

výpočty pro znak y

$$\text{aritmetický průměr } \bar{y} = \frac{134 + 108 + \dots + 80}{18} = \frac{1842}{18} = 102$$

$$\text{směrodatná odchylka } s_y = \sqrt{\frac{1}{18} [(134 - 102)^2 + \dots + (80 - 102)^2]} = \sqrt{\frac{1}{18} \cdot 5756} \doteq 17,882$$

výpočty pro znaky x i y

$$\text{aritmetický průměr } \bar{xy} = \frac{1,00 \cdot 134 + 1,18 \cdot 108 + \dots + 3,45 \cdot 80}{18} = \frac{3895,83}{18} = 216,435$$

Korelační koeficient:

$$r_{xy} = \frac{216,435 - 2,23 \cdot 102}{17,882 \cdot 0,774} = \frac{-11,025}{13,84} \doteq -0,7966$$

Závislost průměrné známky a hodnoty IQ je velká a nepřímá. To znamená, čím vyšší je inteligenční kvocient, tím menší je průměrný prospěch žáka.

4. Rychlost výrobní linky (znak x) se nastavuje pro každou směnu podle potřeby na hodnotu 1 nebo 1,5 nebo 2. Počet poruch za směnu (znak y) se pohybuje od 0 do 3. Máme záznamy za 365 směn (viz. tabulka). Vypočítejte koeficient korelace mezi oběma znaky.

x \ y	0	1	2	3
1	45	20	8	0
1,5	38	62	32	14
2	10	22	58	56

výpočty pro znak x

x_i	1	1,5	2	Σ
n_i	73	146	146	365
$x_i \cdot n_i$	73	219	292	584
$(x_i - \bar{x})^2$	0,36	0,01	0,16	---
$(x_i - \bar{x})^2 \cdot n_i$	26,28	1,46	23,36	51,1

$$\text{aritmetický průměr } \bar{x} = \frac{584}{365} = 1,6$$

směrodatná odchylka

$$s_x = \sqrt{\frac{1}{365} \cdot 51,1} = \sqrt{0,14} = 0,374$$

výpočty pro znak y

x_i	0	1	2	3	Σ
n_i	93	104	98	70	365
$x_i \cdot n_i$	0	104	196	210	510
$(x_i - \bar{y})^2$	1,9516	0,158	0,364	2,5696	---
$(x_i - \bar{y})^2 \cdot n_i$	181,499	16,432	35,672	179,872	413,475

$$\text{aritmetický průměr } \bar{y} = \frac{510}{365} = 1,397$$

směrodatná odchylka

$$s_y = \sqrt{\frac{1}{365} \cdot 413,475} = \sqrt{1,1328} = 1,06$$

výpočty pro znaky x i y

$$\Sigma x_i \cdot y_i = 0 \cdot (1 \cdot 45 + 1,5 \cdot 38 + 2 \cdot 10) + 1 \cdot (1 \cdot 20 + 1,5 \cdot 62 + 2 \cdot 22) + 2 \cdot (1 \cdot 8 + 1,5 \cdot 32 + 2 \cdot 58) + 3 \cdot (1 \cdot 0 + 1,5 \cdot 14 + 2 \cdot 56) = 0 + 157 + 344 + 399 = 900$$

$$\frac{1}{n} \sum_{i=1}^n x_i y_i = \frac{1}{365} \cdot 900 = 2,4658$$

korelační koeficient $r_{xy} = \frac{2,4658 - 1,6 \cdot 1,397}{0,374 \cdot 1,06} = \frac{0,2306}{0,39644} = 0,582$

 Závislost znaků x a y je přímoúměrná.

5. Na 20 pracovištích byla zkoumána vzájemná závislost mezi denním počtem dělníků na pracovišti a denním výkonem pracoviště v ks (viz. tabulka). Vypočtete koeficient korelace mezi oběma znaky.

Výkon v ks	Počet pracovišť s počtem dělníků		
	6	8	10
4	2	3	-
5	1	4	2
6	-	3	5

výpočty pro znak x (výkon)

x_i	4	5	6	Σ
n_i	5	7	8	20
$x_i \cdot n_i$	20	35	48	103
$(x_i - \bar{x})^2$	1,3225	0,0225	0,7225	---
$(x_i - \bar{x})^2 \cdot n_i$	6,6125	0,1575	5,78	12,55

výpočty pro znak y (počet dělníků)

y_i	6	8	10	Σ
n_i	3	10	7	20
$y_i \cdot n_i$	18	80	70	168
$(y_i - \bar{y})^2$	5,76	0,16	2,56	---
$(y_i - \bar{y})^2 \cdot n_i$	17,28	1,6	17,92	36,8

výpočty pro znaky x i y

$$\Sigma x_i \cdot y_i = 4 \cdot (2 \cdot 6 + 3 \cdot 8) + 5 \cdot (1 \cdot 6 + 4 \cdot 8 + 2 \cdot 10) + 6 \cdot (8 \cdot 3 + 10 \cdot 5) = 144 + 290 + 444 = 878$$

korelační koeficient $r_{xy} = \frac{43,9 - 5,15 \cdot 8,4}{0,79215 \cdot 1,35647} = \frac{0,64}{1,07453} = 0,596$

Závislost počtu dělníků a vyrobených výrobků je podle očekávání přímoúměrná.

6. V 8 dílnách různě technicky i personálně vybavených byla v rámci celého výrobního provozu zjištěna následující data (viz. tabulka). Vypočtete závislost mezi výrobou přístroje typu Axa a náklady na jeho výrobu.

dílna	1	2	3	4	5	6	7	8
výroba ks	5	3	7	4	6	5	3	7
celkové náklady v tis. Kč	7	4	8	6	8	7	3	5

výroba ks	3	4	5	6	7	Σ
n_i	2	1	2	1	2	8
$x_i \cdot n_i$	6	4	10	6	14	40
$(x_i - \bar{x})^2$	4	1	0	1	4	---
$(x_i - \bar{x})^2 \cdot n_i$	8	1	0	1	8	18

aritmetický průměr $\bar{x} = \frac{40}{8} = 5$

směrodatná odchylka

$$s_x = \sqrt{\frac{1}{8} \cdot 18} = \sqrt{2,25} = 1,5$$

celkové náklady v tis. Kč	3	4	5	6	7	8	Σ
n_i	1	1	1	1	2	2	8
$y_i \cdot n_i$	3	4	5	6	14	16	48
$(y_i - \bar{y})^2$	9	4	1	0	1	4	---
$(y_i - \bar{y})^2 \cdot n_i$	9	4	1	0	2	8	24

aritmetický průměr $\bar{y} = \frac{48}{8} = 6$

směrodatná odchylka

$$s_x = \sqrt{\frac{1}{8} \cdot 24} = \sqrt{3} = 1,732$$

výpočty pro znaky x i y

$$\Sigma x_i \cdot y_i = 5 \cdot 7 + 3 \cdot 4 + 7 \cdot 8 + 4 \cdot 6 + 6 \cdot 8 + 5 \cdot 7 + 3 \cdot 3 + 7 \cdot 5 = 35 + 12 + 56 + 24 + 48 + 35 + 9 + 35 = 254$$

$$\frac{1}{n} \sum x_i y_i = \frac{1}{8} \cdot 254 = 31,75$$

korelační koeficient $r_{xy} = \frac{31,75 - 5 \cdot 6}{1,5 \cdot 1,732} = \frac{1,75}{2,598} = 0,674$

Závislost počtu výrobků a celkových nákladů je podle očekávání velká a přímoúměrná.

7. U 100 dělníků byla zkoumána závislost mezi měsíční produktivitou práce v tis. Kč a délkou praxe v letech. K dispozici jsou tyto roztríděné informace, které byly uspořádány pro přehlednost do této tabulky. Vypočítejte vzájemnou závislost těchto znaků.

produktivita práce v tis.Kč	Počet dělníků s délkou praxe v dokončených letech					
	- 5	6 - 10	11 - 15	16 - 20	21 - 25	26 - 30
201 - 300	6	2	-	-	1	1
301 - 400	1	11	3	-	-	2
401 - 500	-	3	18	4	4	3
501 - 600	-	1	6	16	6	2
601 - 700	-	-	2	5	3	-

produktivita	250	350	450	550	650	Σ
n_i	10	17	32	31	10	100
$x_i \cdot n_i$	2500	5950	14400	17050	6500	46400
$(x_i - \bar{x})^2$	45796	12996	196	7396	34596	---
$(x_i - \bar{x})^2 \cdot n_i$	457960	220932	6272	229276	345960	1260400

aritmetický průměr

$$\bar{x} = \frac{46400}{100} = 464$$

směrodatná odchylka

$$s_x = \sqrt{\frac{1}{100} \cdot 1260400} = \sqrt{12604} = 112,27$$

počet dělníků	3	8	13	18	23	28	Σ
n_i	7	17	29	25	14	8	100
$y_i \cdot n_i$	21	136	377	450	322	224	1530
$(y_i - \bar{y})^2$	151,29	53,29	5,29	7,29	59,29	161,29	---
$(y_i - \bar{y})^2 \cdot n_i$	1059,03	905,93	153,41	182,25	830,06	1290,32	4421

výpočty pro znaky x i y

$$\Sigma x_i \cdot y_i = 250(3 \cdot 6 + 8 \cdot 2 + 23 \cdot 1 + 28 \cdot 1) + 350(3 \cdot 1 + 8 \cdot 11 + 13 \cdot 3 + 28 \cdot 2) + 450(8 \cdot 3 + 13 \cdot 18 + 18 \cdot 4 + 23 \cdot 4 + 28 \cdot 3) + 550(8 \cdot 1 + 13 \cdot 6 + 18 \cdot 16 + 23 \cdot 6 + 28 \cdot 2) + 650(13 \cdot 2 + 18 \cdot 5 + 23 \cdot 3) = 21250 + 65100 + 227700 + 312400 = \underline{\underline{746700}}$$

aritmetický průměr $\bar{y} = \frac{1530}{100} = 15,3$

směrodatná odchylka

$$s_x = \sqrt{\frac{1}{100} \cdot 4421} = \sqrt{44,21} = 6,649$$

korelační koeficient $r_{xy} = \frac{7467 - 464 \cdot 15,3}{112,27 \cdot 6,649} = \frac{367,8}{746,48323} = 0,493$

$$\frac{1}{n} \sum x_i y_i = \frac{1}{100} \cdot 746700 = 7467$$

Produktivita práce v tis.Kč souvisí s délkou praxe středně a přímoměrně.