5 NÁHODNÝ VEKTOR

Náhodným vektorem rozumíme sloupcový vektor složený z náhodných veličin *X* = (X_1, X_2, ... X_n), který je charakterizován **sdruženou (simultánní) distribuční funkcí.**

$$F(x,y) = P(X < \text{x}; Y < \text{y})$$

náhodný vektor s diskrétním rozdělením: $F(x,y) = \sum_{(i,j)\ x_i<x, y_j<y} P(X = x_i, Y = y_j)$

náhodném vektoru se spojitým rozdělením: $F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x^1, y^1)dx^1dy^1$

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$$

Ze sdruženého rozdělení náhodného vektoru můžeme snadno najít **marginální rozdělení pravděpodobnosti** jednotlivých náhodných veličin, z nichž je vektor sestaven.

Marginální distribuční funkce dvousložkového náhodného vektoru definujeme takto:

$$F_X(x) = P(X < x) = \lim_{y\to\infty} F(x,y)$$
$$F_Y(y) = P(Y < y) = \lim_{x\to\infty} F(x,y)$$

náhodný vektor s diskrétním rozdělením - marginální pravděpodobnosti:

$$P_X(x) = \sum_{y_{jj}} P(X = x, Y = y_j)$$
$$P_Y(y) = \sum_{x_i} P(X = x_i, Y = y)$$

náhodném vektoru se spojitým rozdělením - marginální hustoty pravděpodobnosti:

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy$$
$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)dx$$

Podmíněné rozdělení pak chápeme jako podíl sdruženého a marginálního rozdělení pravděpodobnosti (má-li tento podíl smysl), v souladu s definicí podmíněné pravděpodobnosti.

náhodný vektor s diskrétním rozdělením - podmíněná pravděpodobnostní funkce:

$$P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P_Y(Y = y)}, \qquad pro\ P_Y(Y = y) \neq 0$$

náhodný vektor se spojitým rozdělením - podmíněná hustota pravděpodobnosti:

$$f(x|y) = \frac{f(x,y)}{f_Y(y)}, \qquad pro\ f_Y(y) \neq 0$$

Nezávislost náhodných veličin se projevuje tím, že jejich sdružená distribuční funkce (sdružená pravděpodobnostní funkce, resp. sdružená hustota pravděpodobnosti) se dá matematicky vyjádřit jako součin marginálních distribučních funkcí (marginálních pravděpodobnosti, resp. marginálních hustot pravděpodobnosti) jednotlivých náhodných veličin.

Platí, že složky X, Y náhodného vektoru jsou nezávislé právě když platí:

náhodný vektor s diskrétním rozdělením: $P(X = x_i, Y = y_j) = P_X(X = x_i).P_Y(Y = y_j)$

náhodný vektor se spojitým rozdělením: $f(x,y) = f_X(x).f_Y(y)$

Mezi nejvýznamnější smíšené momenty náhodného vektoru patří **kovariance**.

$$Cov(X,Y) = \mu_{11} = E[(X - EX)(Y - EY)]$$

V praxi se čato setkáváme s reprezentací centrálních momentů 2.řádu ve formě tzv. kovarianční matice:

$$\begin{pmatrix} DX & Cov(X,Y) \\ Cov(X,Y) & DY \end{pmatrix}$$

Mírou lineární závislosti je **korelační koeficient**.

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{DX.DY}} \qquad DX, DY \neq 0$$

5.1. Představme si, že budeme třikrát opakovat pokus u nějž známe pravděpodobnost úspěchu (např. hody mincí, p = 0,5). Zvolme tyto náhodné veličiny:

Y ... počet pokusů do prvního úspěchu
Z ... počet po sobě jdoucích úspěchů

Náhodný vektor X = (Y, Z).

a) Určete marginální pravděpodobnostní funkce $P_Y(y)$, $P_Z(z)$
b) Sestavte sdruženou pravděpodobnostní funkci P(Y=y, Z=z)
c) Určete, zda jsou náhodné veličiny Y, Z nezávislé.
d) Určete střední hodnoty a rozptyly složek Y,Z
e) Určete kovarianční matici Y, Z
f) Určete jednoduchý korelační koeficient
g) Určete podmíněné pravděpodobnostní funkce a P(Y=y | Z=z), P(Z=z | Y=y)

Řešení:

Vypišme si všechny možné kombinace, k nimž by mohlo dojít (S - úspěch, F - neúspěch):
{ FFF; SFS; SSF; FSS; FSF; FFS; SFF; SSS }

A uvažujme, že pravděpodobnost úspěchu P(S) = p, pravděpodobnost neúspěchu P(F) = 1-p.

Jedná se o diskrétní dvourozměrný náhodný vektor, přičemž:

složka Y může nabývat hodnot: 0, 1, 2, 3
složka Z může nabývat hodnot: 0, 1, 2, 3

Pojmenujme si všechny elementární jevy základního prostoru a určeme pravděpodobnost jejich výskytu (pro výpočet jednotlivých pravděpodobností využijeme poznatku, že jevy F a S jsou nezávislé).

A1 ... FFF $P(A1) = (1-p)^3 = 0{,}125$
A2 ... SFS $P(A2) = p^2.(1-p) = 0{,}125$
A3 ... SSF $P(A3) = p^2.(1-p) = 0{,}125$
A4 ... FSS $P(A4) = p^2.(1-p) = 0{,}125$
A5 ... FSF $P(A5) = p.(1-p)^2 = 0{,}125$
A6 ... FFS $P(A6) = p.(1-p)^2 = 0{,}125$
A7 ... SFF $P(A7) = p.(1-p)^2 = 0{,}125$
A8 ... SSS $P(A8) = p^3 = 0{,}125$

ada) Zapišme si nyní do pomocných tabulek, které jevy vyhovují daným hodnotám náhodných veličin Y a Z.

Y ... počet pokusů do prvního úspěchu			
0	1	2	3
A2, A3, A7, A8	A4, A5	A6	A1

Z ... počet po sobě jdoucích úspěchů			
0	1	2	3
A1	A2, A5, A6, A7	A3, A4	A8

protože jevy A1, ... , A8 jsou neslučitelné, můžeme marginální pravděpodobnostní funkce jednoduše určit.

Např. $P_Y(0) = P(A2) + P(A3) + P(A7) + P(A8) = 0{,}125 + 0{,}125 + 0{,}125 + 0{,}125 = 0{,}500$

Y ... počet pokusů do prvního úspěchu			
$P_Y(0)$	$P_Y(1)$	$P_Y(2)$	$P_Y(3)$
0.500	0.250	0.125	0.125

Z ... počet po sobě jdoucích úspěchů			
$P_Z(0)$	$P_Z(1)$	$P_Z(2)$	$P_Z(3)$
0.125	0.500	0.250	0.125

V našem případě (máme určovat zároveň sdruženou pravděpodobnostní funkci) by bylo rychlejší pro určeni marginálních pravděpodobnostních funkcí využít korelační tabulku, kterou budeme vytvářet pro zápis sdružené pravděpodobnosti.

adb) Zkonstruujeme korelační tabulku. (nejdříve si do ní vypíšeme jevy, které vyhovují příslušným podmínkám a poté na základě jejich neslučitelnosti určíme pravděpodobnosti výskytu příslušných skupin jevů)

		Z			
		0	1	2	3
Y	0	-	A2, A7	A3	A8
	1	-	A5	A4	-
	2	-	A6	-	-
	3	A8	-	-	-

Tabulka sdružené pravděpodobnostní funkce

		Z			
		0	1	2	3
Y	0	0	0,250	0,125	0,125
	1	0	0,125	0,125	0
	2	0	0,125	0	0
	3	0,125	0	0	0

Např. P(Y = 0, Z = 2) = 0,125; P(Y = 0, Z = 1) = 0,250

Chceme-li získat korelační tabulku v klasickém tvaru, tj. včetně marginálních pravděpodobnosti, stačí sečíst příslušné řádky (sloupce).

		Z				
		0	1	2	3	$P_Y(y)$
Y	0	0	0,250	0,125	0,125	0,500
	1	0	0,125	0,125	0	0,250
	2	0	0,125	0	0	0,125
	3	0,125	0	0	0	0,125
	$P_Z(z)$	0,125	0,500	0,250	0,125	1

Pro srovnání si porovnejte takto získané marginální pravděpodobnosti s marginálními pravděpodobnostmi získanými v ada)

adc) Pro náhodný vektor s diskrétním rozdělením platí, že složky Y, Z náhodného vektoru jsou nezávislé právě když platí:

$$P(Y = y_i, Z = z_j) = P_Y(Y = y_i).P_Z(Z = z_j)$$

Tento předpoklad v našem případě splněn není. (např. $P(\acute{Y} = 2, Z = 1) \neq P_Y(2).P_Z(1)$; $0{,}125 \neq 0{,}125.0{,}500$). Z toho plyne, že **náhodné veličiny Y, Z nejsou nezávislé**.

add) Střední hodnoty a rozptyly získáme z definičních vztahů pomocí marginálních pravděpodobnostních funkcí:

$$\underline{\underline{EY}} = \sum_{i=1}^{4} y_i . P_Y(y_i) = 0.0{,}500 + 1.0{,}250 + 2.0{,}125 + 3.0{,}125 = \frac{7}{8} = \underline{\underline{0{,}875}}$$

$$EY^2 = \sum_{i=1}^{4} y_i^2 . P_Y(y_i) = 0^2.0{,}500 + 1^2.0{,}250 + 2^2.0{,}125 + 3^2.0{,}125 = 1{,}875$$

$$\underline{\underline{DY}} = EY^2 - (EY)^2 = 1{,}875 - (0{,}875)^2 = \frac{71}{64} \cong \underline{\underline{1{,}109}}$$

$$\underline{\underline{EZ}} = \sum_{i=1}^{4} z_i . P_Z(z_i) = 0.0{,}125 + 1.0{,}500 + 2.0{,}250 + 3.0{,}125 = \frac{11}{8} = \underline{\underline{1{,}375}}$$

$$EZ^2 = \sum_{i=1}^{4} z_i^2 . P_Z(z_i) = 0^2.0{,}125 + 1^2.0{,}500 + 2^2.0{,}250 + 3^2.0{,}125 = 2{,}625$$

$$\underline{\underline{DZ}} = EZ^2 - (EZ)^2 = 2{,}625 - (1{,}375)^2 = \frac{47}{64} \cong \underline{\underline{0{,}734}}$$

ade) Kovarianční matice má obecný tvar:

$$\begin{pmatrix} DY & Cov(Y,Z) \\ Cov(Y,Z) & DZ \end{pmatrix}$$

Pro její zápis musíme určit kovarianci.

$$\underline{\underline{Cov(Y,Z)}} = E[(Y-EY)(Z-EZ)] = \sum_{i,j} \left(y_i - \frac{7}{8}\right)\left(z_j - \frac{11}{8}\right) P(Y = y_i, Z = z_j) =$$
$$= \left(0-\frac{7}{8}\right)\left(1-\frac{11}{8}\right).0{,}250 + \left(0-\frac{7}{8}\right)\left(2-\frac{11}{8}\right).0{,}125 + \left(0-\frac{7}{8}\right)\left(3-\frac{11}{8}\right).0{,}125 +$$
$$+ \left(1-\frac{7}{8}\right)\left(1-\frac{11}{8}\right).0{,}125 + \left(1-\frac{7}{8}\right)\left(2-\frac{11}{8}\right).0{,}125 + \left(2-\frac{7}{8}\right)\left(1-\frac{11}{8}\right).0{,}125 +$$
$$+ \left(3-\frac{7}{8}\right)\left(0-\frac{11}{8}\right).0{,}125 = -\frac{296}{512} \cong \underline{\underline{-0{,}578}}$$

Kovarianční matice má tvar:

$$\begin{pmatrix} 1{,}109 & -0{,}578 \\ -0{,}578 & 0{,}734 \end{pmatrix}$$

adf) Jednoduchý korelační koeficient určíme z definičního vztahu:

$$\underline{\underline{\rho_{X,Y}}} = \frac{Cov(X,Y)}{\sqrt{DX.DY}} = \frac{-0{,}578}{\sqrt{1{,}109.0{,}734}} \cong \underline{\underline{-0{,}641}}$$

Na základě této hodnoty korelačního koeficientu můžeme říci, že mezi náhodnými veličinami Y a Z existuje **středně silná negativní korelace**, tj. že pravděpodobně s růstem Y bude Z klesat (lineárně).

adg) Podmíněné pravděpodobnosti budeme opět zapisovat do tabulky a jejich hodnoty určíme z definice:

$$P(Y = y|Z = z) = \frac{P(Y = y, Z = z)}{P_Z(z)}$$

Tabulka podmíněné pravděpodobnostní funkce P(Y = y| Z = z)

		Z			
		0	1	2	3
Y	0	0/0,125	0,250/0,500	0,125/0,250	0,125/0,125
	1	0/0,125	0,125/0,500	0,125/0,250	0/0,125
	2	0/0,125	0,125/0,500	0/0,250	0/0,125
	3	0,125/0,125	0/0,500	0/0,250	0/0,125

		Z			
		0	1	2	3
Y	0	0	0,500	0,500	1,000
	1	0	0,250	0,500	0
	2	0	0,250	0	0
	3	1,000	0	0	0

Např. P(Y = 2| Z = 1) = 0,250

$$P(Z = y|Y = y) = \frac{P(Y = y, Z = z)}{P_Y(y)}$$

Tabulka podmíněné pravděpodobnostní funkce P(Z = z| Y = y)

		Z			
		0	1	2	3
Y	0	0/0,500	0,250/0,500	0,125/0,500	0,125/0,500
	1	0/0,250	0,125/0,250	0,125/0,250	0/0,250
	2	0/0,125	0,125/0,125	0/0,125	0/0,125
	3	0,125/0,125	0/0,125	0/0,125	0/0,125

		Z			
		0	1	2	3
Y	0	0	0,500	0,250	0,250
	1	0	0,500	0,500	0
	2	0	1,000	0	0
	3	1,000	0	0	0

Např. P(Z = 2| Y= 1) = 0,500

5.2. Studenti z jedné studijní skupiny byli na zkoušce z matematiky a fyziky s těmito výsledky (první hodnota v uspořádané dvojici označuje výsledek studenta z matematiky, druhá z fyziky):

(1,1), (1,2), (1,3), (2,2), (2,3), (2,3), (3,2), (3,2), (3,3), (3,3), (3,3), (3,3), (3,3), (3,4), (3,4), (4,3), (4,3), (4,4), (4,4), (4,4).

a) Vytvořte pravděpodobnostní tabulku náhodného vektoru, jehož složka X bude znamenat výsledky u zkoušky z matematiky a složka Y bude znamenat výsledky u zkoušky z fyziky
b) Určete jeho marginální pravděpodobnostní funkce $P_X(x)$, $P_Y(y)$
c) Určete jeho distribuční funkci F(x,y)
d) Zjistěte jeho podmíněné pravděpodobnosti P(X=x/Y=y)

Řešení:
ada)

Tabulka sdružených četností:

		Y			
		1	**2**	**3**	**4**
	1	1	1	1	0
	2	0	1	2	0
X	**3**	0	2	5	2
	4	0	0	2	3

Celkem: 20

Tabulka sdružených pravděpodobností:

		Y			
		1	**2**	**3**	**4**
	1	0,05	0,05	0,05	0
	2	0	0,05	0,10	0
X	**3**	0	0,10	0,25	0,10
	4	0	0	0,10	0,15

Hodnoty v prvním řádku a prvním sloupci jsou hodnoty, kterých mohou nabývat náhodné veličiny *X, Y*. Ostatní čísla v tabulce jsou pravděpodobnosti výskytu všech možných dvojic.

např. $P(1;1) = \frac{1}{20} = 0{,}05$

adb)

		Y				
		1	**2**	**3**	**4**	**$P_X(x_i)$**
	1	0,05	0,05	0,05	0	0,15
	2	0	0,05	0,10	0	0,15
X	**3**	0	0,10	0,25	0,10	0,45
	4	0	0	0,10	0,15	0,25
	$P_Y(y_i)$	0,05	0,20	0,50	0,25	**1,00**

Hodnoty marginální pravděpodobnostní funkce $P_X(x_i)$ jsou vždy součty všech pravděpodobností v daném řádku.

např.:$P_X(3) = 0 + 0,1 + 0,25 + 0,1 = 0,45$.

Obdobně nalezneme ve sloupcích hodnoty $P_Y(y_i)$ Zvýrazněné číslo musí být vždy rovno jedné, je to součet všech hodnot $P_X(x_i)$ nebo $P_Y(y_i)$, tedy vlastně součet všech sdružených pravděpodobností náhodného vektoru.

adc)

F(x,y):

		Y 1	2	3	4	5
	1	0	0	0	0	0
	2	0	0,05	0,10	0,15	0,15
X	**3**	0	0,05	0,15	0,30	0,30
	4	0	0,05	0,25	0,65	0,75
	5	0	0,05	0,25	0,75	1,00

postup při výpočtu:

např.: $F(3,3) = P(X<3,Y<3) = P(1,1) + P(1,2) + P(2,1) + P(2,2) = 0,15$

Všimněte si, že hodnoty v posledním sloupci odpovídají hodnotám marginální distribuční funkce $F_X(x)$ a hodnoty v posledním řádku hodnotám $F_Y(y)$

add)

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P_Y(Y = y)}$$

		Y 1	2	3	4
	1	1,00	0,25	0,10	0
	2	0	0,25	0,20	0
X	**3**	0	0,50	0,50	0,40
	4	0	0	0,20	0,60

např.:

$$P(X = 3|Y = 3) = \frac{P(X = 3, Y = 3)}{P_Y(Y = 3)} = \frac{0,25}{0,50} = 0,50$$

5.3. Sdružená hustota pravděpodobnosti dvousložkového náhodného vektoru je definována jako:

$$f(x,y)=\begin{matrix} x+y & pro\ (x,y)\in\langle 0;1\rangle\times\langle 0;1\rangle \\ 0 & jinde \end{matrix}$$

Určete:
a) Marginální hustoty pravděpodobnosti $f_X(x)$, $f_Y(y)$
b) Marginální distribuční funkce $F_X(x)$, $F_Y(y)$
c) Střední hodnoty a rozptyly složek X, Y
d) Hodnotu jednoduchého korelačního koeficientu, výsledek dejte do souvislosti s mírou lineární závislosti

Řešení:

ada) Jde o spojitý náhodný vektor, proto:

$$f_X(x)=\begin{matrix} \int_0^1 (x+y)dy=\left[xy+\frac{y^2}{2}\right]_0^1=x+\frac{1}{2} & pro\ x\in\langle 0;1\rangle \\ 0 & jinde \end{matrix}$$

Ze symetrie sdružené pravděpodobnostní funkce vyplývá i obdobný tvar $f_Y(y)$.

$$f_Y(y)=\begin{matrix} y+\frac{1}{2} & pro\ y\in\langle 0;1\rangle \\ 0 & jinde \end{matrix}$$

adb) Marginální distribuční funkce jednotlivých složek určíme z marginálních hustot pravděpodobnosti:

$$F_X(x)=\begin{matrix} \int_{-\infty}^{x} 0dt=0 & pro\ x\in(-\infty;0) \\ \int_{-\infty}^{0} 0dt+\int_0^x\left(t+\frac{1}{2}\right)dt=0+\left[\frac{t^2}{2}+\frac{t}{2}\right]_0^x=\frac{1}{2}x(x+1) & pro\ x\in\langle 0;1\rangle \\ \int_{-\infty}^{0} 0dt+\int_0^1\left(t+\frac{1}{2}\right)dt+\int_0^x 0dt=0+1+0=1 & pro\ x\in(1;\infty) \end{matrix}$$

ze symetrie f(x,y) můžeme opět odvodit:

$$F_Y(y)=\begin{matrix} 0 & pro\ y\in(-\infty;0) \\ \frac{1}{2}y(y+1) & pro\ y\in\langle 0;1\rangle \\ 1 & pro\ y\in(1;\infty) \end{matrix}$$

adc) Střední hodnoty a rozptyly jednotlivých složek určíme pomocí marginálních hustot pravděpodobnosti, na základě znalosti definičních vztahů pro oba momenty:

$$\underline{\underline{EX}} = \int_{-\infty}^{\infty} x.f_X(x)dx = \int_0^1 x(x+\frac{1}{2})dx = \left[\frac{x^3}{3}+\frac{x^2}{4}\right]_0^1 = \underline{\underline{\frac{7}{12}}}$$

$$EX^2 = \int_{-\infty}^{\infty} x^2.f_X(x)dx = \int_0^1 x^2(x+\frac{1}{2})dx = \left[\frac{x^4}{4}+\frac{x^3}{6}\right]_0^1 = \frac{5}{12}$$

$$\underline{\underline{DX}} = EX^2-(EX)^2 = \frac{5}{12}-\left(\frac{7}{12}\right)^2 = \frac{60-49}{144} = \underline{\underline{\frac{11}{144}}}$$

Opět využijeme symetrie sdružené hustoty pravděpodobnosti f(x,y) a můžeme tvrdit, že:

$$\underline{\underline{EY = \frac{7}{12}}};\quad \underline{\underline{\mathrm{DY} = \frac{11}{144}}}$$

add) Pro výpočet jednoduchého korelačního koeficientu potřebujeme znát hodnotu kovariance a proto začneme jejím výpočtem:

$$\underline{\underline{Cov(X,Y)}} = E[(X-EX)(Y-EY)] = \int_0^1\int_0^1\left(x-\frac{7}{12}\right).\left(y-\frac{7}{12}\right).(x+y)dxdy =$$

$$= \int_0^1\int_0^1\left[\left(x^2-\frac{7}{12}x\right).\left(y-\frac{7}{12}\right)+\left(x-\frac{7}{12}\right)\left(y^2-\frac{7}{12}y\right)\right]dxdy =$$

$$= \int_0^1\left[\left(\frac{x^3}{3}-\frac{7x^2}{24}\right).\left(y-\frac{7}{12}\right)+\left(\frac{x^2}{2}-\frac{7x}{12}\right).\left(y^2-\frac{7}{12}y\right)\right]_0^1 dy =$$

$$= \int_0^1\left[\frac{1}{24}\left(y-\frac{7}{12}\right)-\frac{1}{12}.\left(y^2-\frac{7}{12}y\right)\right]dy = \left[\frac{1}{24}\left(\frac{y^2}{2}-\frac{7y}{12}-\frac{2y^3}{3}+\frac{7y^2}{12}\right)\right]_0^1 = \underline{\underline{\frac{-1}{144}}}$$

Dále již stačí jen dosadit do definičního vztahu pro jednoduchý korelační koeficient:

$$\underline{\underline{\rho_{X,Y}}} = \frac{Cov(X,Y)}{\sqrt{DX.DY}} = \frac{\frac{-1}{144}}{\sqrt{\frac{11}{144}.\frac{11}{144}}} = \underline{\underline{\frac{-1}{11}}} \cong -0{,}091$$

Z velikosti korelačního koeficientu můžeme usuzovat na to, že mezi X a Y pravděpodobně neexistuje lineární závislost, tj, X a Y jsou **nekorelované** náhodné veličiny.

5.4. Vypočtěte střední hodnotu náhodné veličiny *X* náhodného vektoru, který je určen hustotou pravděpodobnosti:

$$f(x,y)=\begin{cases}0{,}5.\sin(x+y) & \text{pro } 0\le x\le \frac{\pi}{2},\, 0\le y\le \frac{\pi}{2}\\ 0 & \text{jinde}\end{cases}$$

Řešení:

$$EX=\int_{-\infty}^{\infty} x\cdot f_X(x)dx,\ \text{kde}\ f_X(x)=\int_{-\infty}^{\infty} f(x,y)dy$$

$$x\in\left\langle 0;\frac{\pi}{2}\right\rangle:$$

$$f_X(x)=\int_{-\infty}^{\infty} f(x,y)dy=\frac{1}{2}\int_0^{\frac{\pi}{2}}\sin(x+y)dy=-\frac{1}{2}\left[\cos(x+y)\right]_0^{\frac{\pi}{2}}=-\frac{1}{2}\left(\cos\left(x+\frac{\pi}{2}\right)-\cos(x)\right)$$

$$=\frac{1}{2}\left(\cos(x)-\cos\left(x+\frac{\pi}{2}\right)\right)$$

$$EX=\int_{-\infty}^{\infty} x\cdot f_X(x)dx=\int_0^{\frac{\pi}{2}} x\cdot\frac{1}{2}\left(\cos(x)-\cos\left(x+\frac{\pi}{2}\right)\right)dx=\frac{1}{2}\cdot\left[\int_0^{\frac{\pi}{2}} x\cdot\left[\cos\left(x+\frac{\pi}{2}\right)-\cos(x)\right]dx\right]$$

Pro vyřešení tohoto integrálu použijeme metodu per partes:

$$\underline{\underline{EX}}=\frac{1}{2}\int_0^{\frac{\pi}{2}} x\cdot\left[\cos\left(x+\frac{\pi}{2}\right)-\cos x\right]dx=\left|\begin{matrix}u=x & v'=\cos\left(x+\frac{\pi}{2}\right)-\cos(x)\\ u'=1 & v=-\sin\left(x+\frac{\pi}{2}\right)+\sin(x)\end{matrix}\right|=$$

$$=\frac{1}{2}\left\{\left[x\cdot\left(-\sin\left(x+\frac{\pi}{2}\right)+\sin(x)\right)\right]_0^{\frac{\pi}{2}}+\int_0^{\frac{\pi}{2}}\left(\sin\left(x+\frac{\pi}{2}\right)-\sin(x)\right)dx\right\}=$$

$$=\frac{1}{2}\left\{\left[x\cdot\left(-\sin\left(x+\frac{\pi}{2}\right)+\sin(x)\right)\right]_0^{\frac{\pi}{2}}+\left[\left(\cos\left(x+\frac{\pi}{2}\right)-\cos(x)\right)\right]_0^{\frac{\pi}{2}}\right\}=$$

$$=\frac{1}{2}\left\{\left[x\cdot\left(-\sin\left(x+\frac{\pi}{2}\right)+\sin(x)\right)+\cos\left(x+\frac{\pi}{2}\right)-\cos(x)\right]_0^{\frac{\pi}{2}}\right\}=$$

$$=\frac{1}{2}\left\{\left(\frac{\pi}{2}\cdot(0+1)+(-1)-0\right)-\left(0\cdot(-1+0)+0-1\right)\right\}=\frac{1}{2}\cdot\frac{\pi}{2}=\underline{\underline{\frac{\pi}{4}}}$$

Podobným způsobem by se daly vypočítat i zbylé číselné charakteristiky: rozptyl, kovariance a koeficient korelace.

Na závěr cvičení si ukážeme jak můžeme využít Statgraphics při zpracování diskrétního dvourozměrného vektoru.

5.5. Studenti z jedné studijní skupiny byli na zkoušce z matematiky a fyziky s těmito výsledky (první hodnota v uspořádané dvojici označuje výsledek studenta z matematiky, druhá z fyziky):

(1,1), (1,2), (1,3), (2,2), (2,3), (2,3), (3,2), (3,2), (3,3), (3,3), (3,3), (3,3), (3,3), (3,4), (3,4), (4,3), (4,3), (4,4), (4,4), (4,4).

Zvolme tyto náhodné veličiny:

Y ... známka z matematiky
Z ... známka z fyziky

Náhodný vektor X = (Y, Z).

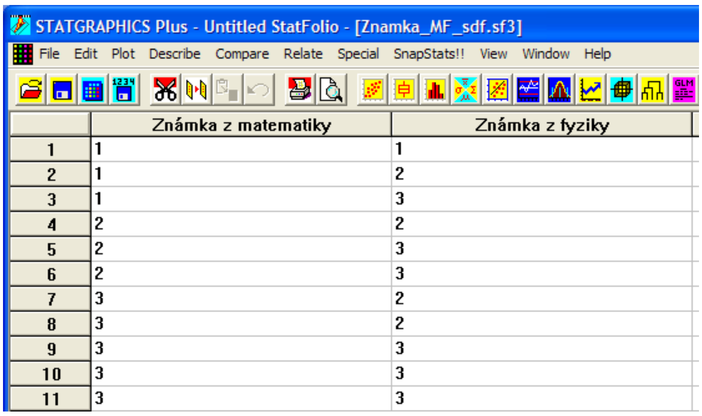
Pomocí Statgraphicsu:

a) Sestavte sdruženou pravděpodobnostní funkci P(Y=y, Z=z)
b) Určete marginální pravděpodobnosti
c) Určete, zda jsou náhodné veličiny Y, Z nezávislé.
d) Určete kovarianční matici Y, Z
e) Určete jednoduchý korelační koeficient

Řešení:

Zpracování dvourozměrného diskrétního náhodného vektoru ve Statgraphicsu zahájíme tím, že do tohoto softwaru zadáme data. Zadání dvourozměrné proměnné provedeme buď v tzv. **standardním datovém formátu** nebo ve formě **tabulky sdružených četností – kontingenční tabulky**.

Pod pojmem **standardní datový formát** si představme klasické zadání dat – definujeme dvě proměnné (známka z matematiky, známka z fyziky) a zadáme všechny kombinace obou proměnných.



	Známka z matematiky	Známka z fyziky
1	1	1
2	1	2
3	1	3
4	2	2
5	2	3
6	2	3
7	3	2
8	3	2
9	3	3
10	3	3
11	3	3

Kontingenční tabulka – tj. v podstatě tabulka sdružených četností.

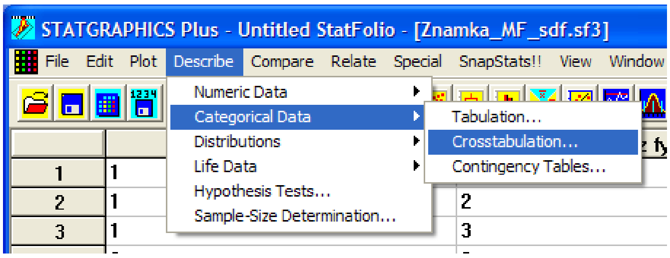
Známka z matematiky, F1, F2, F3 i F4 jsou numerické proměnné. F1 označuje jedničku z fyziky, F2 dvojku z fyziky, apod.

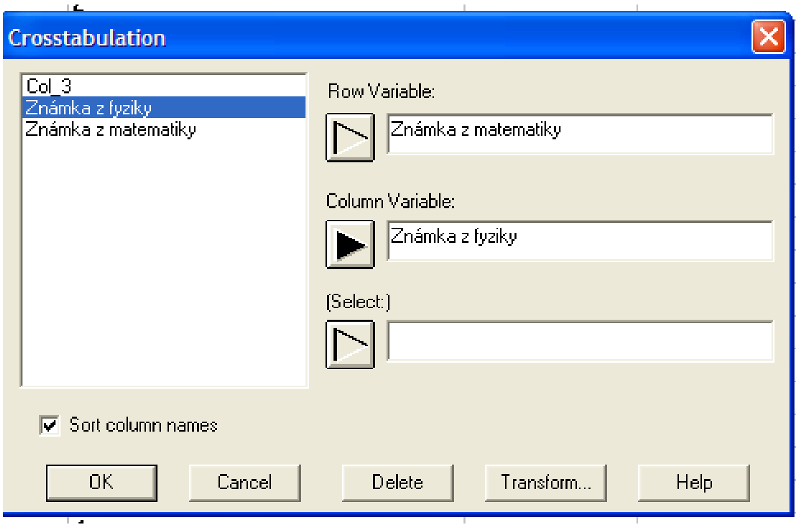
Známka z matematiky	F1	F2	F3	F4
1	1	1	1	0
2	0	1	2	0
3	0	2	5	2
4	0	0	2	3

ada,adb) Sdruženou pravděpodobnostní funkci získáme jako textový výstup při zpracování kategoriální proměnné:

Máme-li data zadána ve **standardním datovém formátu**, pak použijeme:

Menu Describe \ Categorical Data \ Crosstabulation

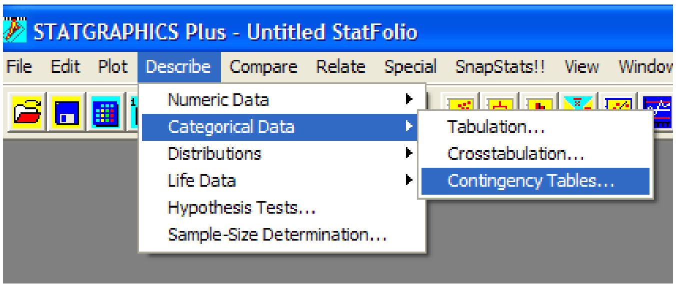


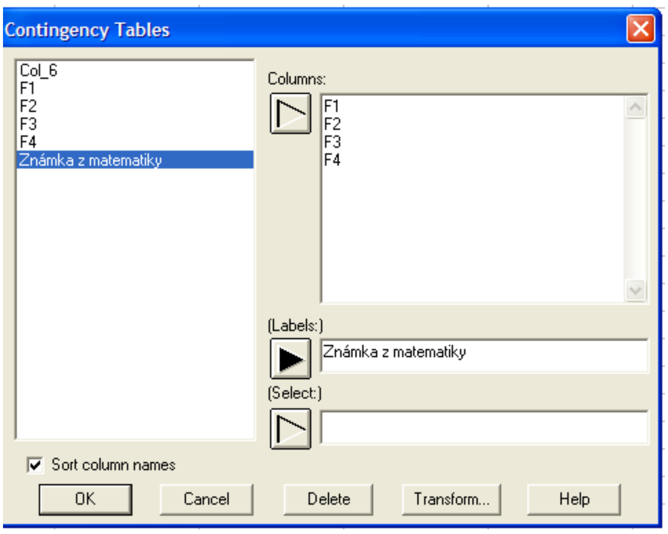


V okně Crosstabulation označíme jednu proměnnou (Známka z matematiky) jako Row Variables a druhou (Známka z fyziky) jako Column Variables. Pokud mezi proměnnými existuje příčinná souvislost, pak za řádkovou proměnnou (Row Variable) volíme proměnnou, která je příčinou změn proměnné, kterou označíme za sloupcovou (Column variable). (např. Množství hnojiva je příčinou změn proměnné Výnosy).

Máme-li data zadána ve formě **kontingenční tabulky**, pak použijeme:

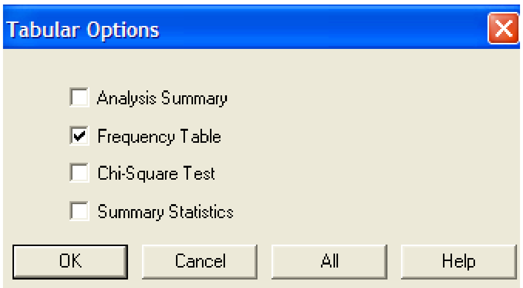
Menu Describe \ Categorical Data \ Contingency Tables





V okně Contingency Tables označíme F1, F2, F3, F4 jako Columns (sloupce) a Známku z matematiky jako Labels.

Dále pokračujeme v obou případech stejně:



Označíme **OK** a pomocí ikony **Tabular Options** (žlutá ikona) si jako požadovaný textový výstu zvolíme tabulku četností – **Frequency Tables**.

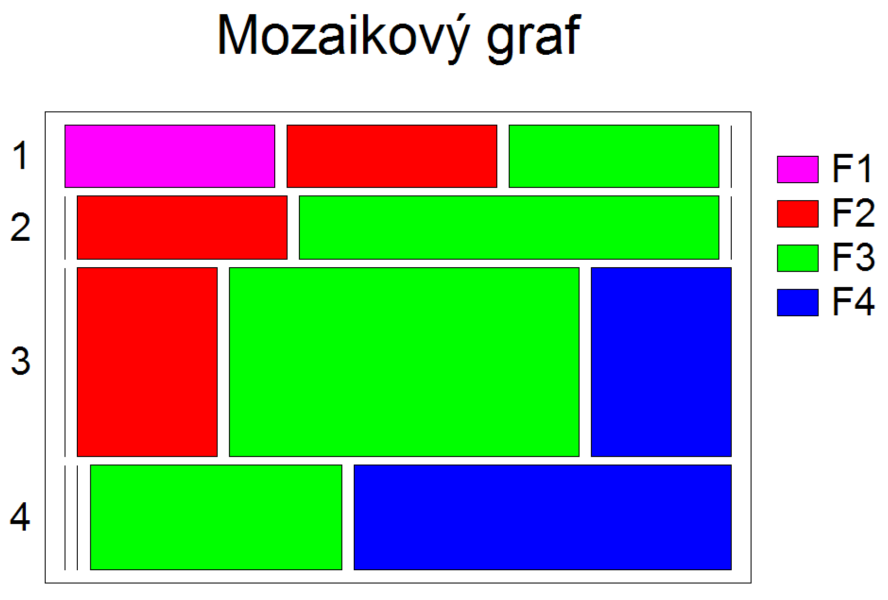
	F1	F2	F3	F4	Row Total
1	1 5,00%	1 5,00%	1 5,00%	0 0,00%	3 15,00%
2	0 0,00%	1 5,00%	2 10,00%	0 0,00%	3 15,00%
3	0 0,00%	2 10,00%	5 25,00%	2 10,00%	9 45,00%
4	0 0,00%	0 0,00%	2 10,00%	3 15,00%	5 25,00%
Column Total	1 5,00%	4 20,00%	10 50,00%	5 25,00%	20 100,00%

Cell contents:
Observed frequency
Percentage of table

Označíme **OK** a jako výstup se nám objeví tabulka, v níž jsou uvedeny jak sdružené četnosti, tak i sdružené pravděpodobnosti (v procentech) a na okrajích tabulky najdeme marginální pravděpodobnosti (%).

adc) Grafickou obdobou kontingenční tabulky je **mozaikový graf**. Tento graf se skládá z obdélníků, jejichž strany jsou úměrné příslušným marginálním relativním četnostem. Statgraphics konstruuje mozaikový graf tak, že na svislou osu vynáší nezávisle proměnnou (příčina) a na vodorovnou osu závisle proměnnou (důsledek). Pokud by byl mozaikový graf v tomto případě tvořen svislými pruhy (jednotlivé obdélníky stejných barev by měly stejné „vodorovné" rozměry), znamenalo by to, že sledované proměnné jsou nezávislé. Obdobné vyhodnocení provedeme v případě, kdy statistický software vynáší nezávisle proměnnou na vodorovnou osu (např. JMP-IN). Pak je v případě nezávislosti proměnných mozaikový graf tvořen vodorovnými pásy.

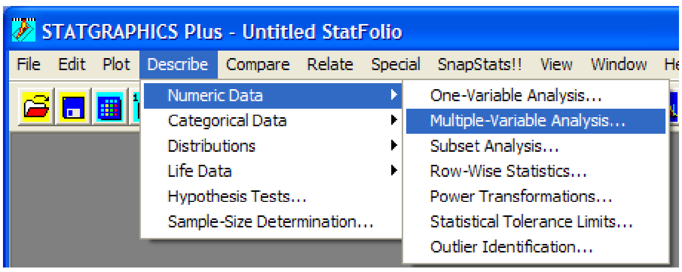
V našem případě nedokážeme určit, která náhodná veličina ovlivňuje kterou a proto nezáleží na tom, kterou budeme vynášet na osu x a kterou na osu y.



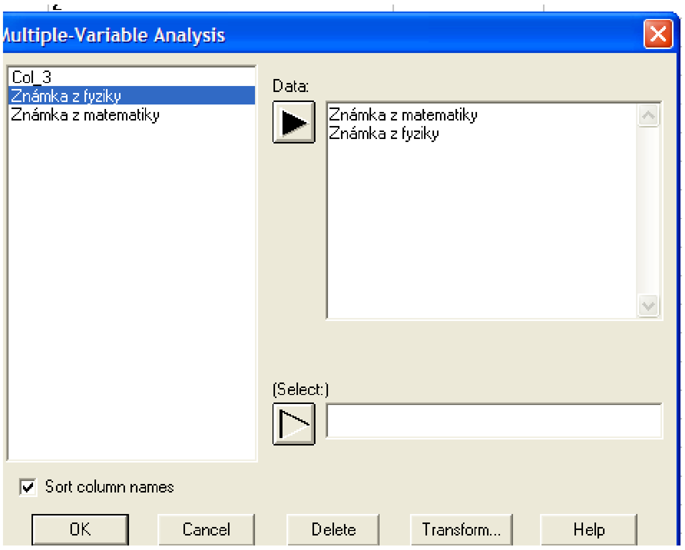
Z grafu je zřejmé, že se jedná o závislé náhodné veličiny. (Toto je pouze závěr explorační (popisné) statistiky).

add) Chceme-li získat **kovarianční matici**, musíme mít dat zadána ve standardním datovém formátu.

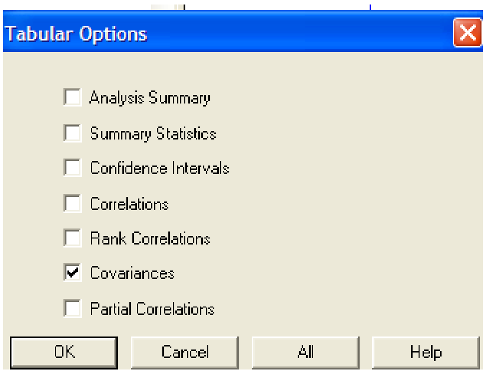
Použijeme menu: **Menu Describe \ Numeric Data \ Multiple – Variable Analysis …**



V okně Multiple – Variable analysis zadáme dané proměnné jako Data a zvolíme OK.



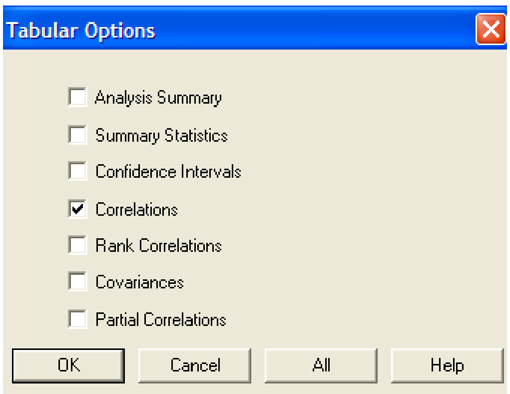
Kovarianční matici získáme zaškrtnutím pole **Covariances** v okně Tabular Options, které se nám objeví po kliknutí na ikonu Tabular Options (žlutá ikona).



Covariances

	Známka z matematiky	Známka z fyziky
Známka z matematiky	1,01053 (20)	0,515789 (20)
Známka z fyziky	0,515789 (20)	0,681579 (20)

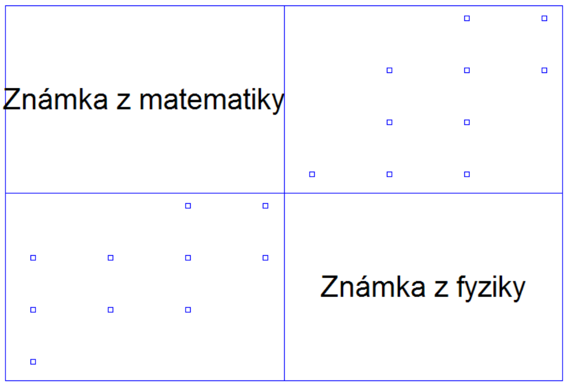
ade) Jednoduchý korelační koeficient získáme obdobně jako kovarianční matici. Pouze v okně Tabular option (žlutá ikona) zaškrtneme pole **Correlations**.



```
Correlations

                      Známka z matematiky Známka z fyziky
------------------------------------------------------------
Známka z matematiky                            0,6215
                                              (    20)
                                               0,0034

Známka z fyziky          0,6215
                        (    20)
                         0,0034
------------------------------------------------------------

Correlation
(Sample Size)
P-Value
```

V textovém výstupu najdeme hodnotu korelačního koeficientu (0,6215), počet hodnot proměnných a hodnotu p-value (budeme se jí zabývat při testování hypotéz), která nám říká zda se korelační koeficient odlišuje od nulové hodnoty natolik, abychom data mohli považovat za lineárně závislá. (Je-li p-value menší než 0,01, pak data považujeme za lineárně závislá.)

Při této analýze získáme rovněž grafický výstup ve formě bodových grafů ukazujících formu závislosti mezi proměnnými.



Z grafu je lineární závislost proměnných zcela zřejmá. Tomu odpovídá i hodnota korelačního koeficientu (0,6215), která ukazuje na silnou pozitivní korelaci. (což je potvrzeno i hodnotou p-value (0,0034 <<< 0,01).