

1. Máme k dispozici 4 spravedlivé šestistěnné kostky a jednu falešnou, na které padne šestka s pravděpodobností 0,5.
- a) Hodíme náhodně vybranou kostkou. Jaká je pravděpodobnost, že padne šestka? (5b)
- b) Na náhodně vybrané kostce padla šestka. Jaká je pravděpodobnost, že tato kostka je falešná? (5b)

$S(F)$... je vybraná spravedlivá (falešná) kostka
 $V6$... padne šestka

$\frac{4}{5}$	S	$\frac{1}{6}$	$V6$
		$\frac{5}{6}$	$\bar{V6}$
$\frac{1}{5}$	F	$\frac{1}{2}$	$V6$
		$\frac{1}{2}$	$\bar{V6}$

$$\begin{aligned}
 a) P(V6) &= \frac{4}{5} \cdot \frac{1}{6} + \frac{1}{5} \cdot \frac{1}{2} = \\
 &= \frac{4}{30} + \frac{3}{30} = \frac{7}{30} \approx \\
 &\approx \underline{\underline{0,233}}
 \end{aligned}$$

$$\begin{aligned}
 b) P(F|V6) &= \frac{P(F \cap V6)}{P(V6)} = \\
 &= \frac{1/5 \cdot 1/2}{7/30} = \\
 &= \frac{3}{7} \approx \underline{\underline{0,429}}
 \end{aligned}$$

2. V určité firmě bylo zjištěno, že na 10 % počítačů je nainstalován nelegální software.
- Určete pravděpodobnostní a distribuční funkci počtu počítačů s nelegálním softwarem mezi dvěma kontrolovanými počítači. (3b)
 - Určete střední hodnotu, směrodatnou odchylku a modus počtu počítačů s nelegálním softwarem mezi dvěma kontrolovanými počítači. (3b)

Uvažujme náhodnou veličinu T , která je modelem doby potřebné ke kontrole a případné „opravě“ (v minutách) dvou kontrolovaných počítačů. Nechť $T = 40 + 60X$.

- Určete střední hodnotu, směrodatnou odchylku a modus doby T . (4b)

X ... počet počítačů s neleg. softwarem (ze 2)

a)

x	0	1	2	Σ
$P(x)$	0,81	0,18	0,01	1,00

$$P(X=0) = 0,9 \cdot 0,9 = 0,81$$

$$P(X=2) = 0,1 \cdot 0,1 = 0,01$$

$$P(X=1) = 1 - P(X=0) - P(X=2) = 0,18$$

x	$(-\infty; 0 >$	$(0; 1 >$	$(1; 2 >$	$(2; \infty)$
$F(x)$	0,00	0,81	0,99	1,00

$$b) \quad \underline{E(X)} = \sum_{(i)} x_i \cdot P(x_i) = 0 \cdot 0,81 + 1 \cdot 0,18 + 2 \cdot 0,01 = \underline{0,20}$$

$$E(X^2) = \sum_{(i)} x_i^2 \cdot P(x_i) = 0^2 \cdot 0,81 + 1^2 \cdot 0,18 + 2^2 \cdot 0,01 = 0,22$$

$$D(X) = E(X^2) - (E(X))^2 = 0,22 - 0,20^2 = 0,18$$

$$\underline{\underline{\sigma(X)}} = \sqrt{D(X)} = \underline{\underline{0,42}}$$

$$\underline{\underline{\hat{x}}} = \underline{\underline{0}}$$

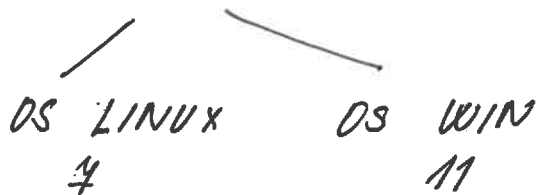
$$c) T = 40 + 60 X$$

$$\underline{E(T)} = 40 + 60 \cdot \underbrace{E(X)}_{0,2} = \underline{52 \text{ min}}$$

$$\underline{\sigma(T)} = 60 \cdot \sigma(X) = \underline{20 \text{ min}}$$

$$\underline{\hat{T}} = \underline{40 \text{ min}}$$

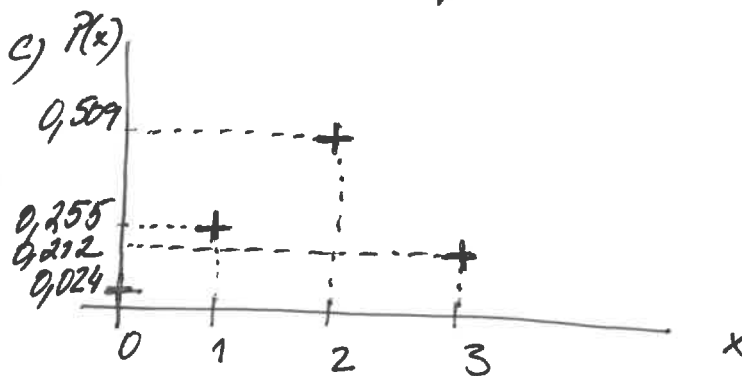
3. V počítačové učebně je na jedenácti počítačích nainstalován pouze operační systém Windows a na zbylých sedmi počítačích pouze operační systém Linux. Zapneme náhodně tři počítače.
- Jaká je pravděpodobnost, že na všech třech počítačích je nainstalován systém Linux. (ze třech zapnutých), na nichž je nainstalován OS Linux. Načrtněte příslušnou pravděpodobnostní funkci. (3b)
 - Určete, jaká je pravděpodobnost, že je na všech třech zapnutých počítačích nainstalován stejný OS. (2b)
 - Načrtněte pravděpodobnostní funkci počtu počítačů s OS Linux (ze tří zapnutých). (5b)



X ... počet počítačů s OS Linux ke třem
 $X \sim H(N=12, M=4, n=3)$

a) $P(X=3) = 0,212$ (dhyper(3, 4, 12, 3))

b) $P(X=3) + P(X=0) = 0,236$
 dhyper(0, 4, 12, 3)



Datový soubor iris.xlsx (R. A. Fisher (1936), <http://home1.vsb.cz/~lit40/DATA/iris.xlsx>). „The use of multiple measurements in taxonomic problems“. Annals of Eugenics. 7(2): 179 (188) obsahuje údaje o 50 rostlinách tří odrůd kosatců (Iris setosa, Iris virginica a Iris versicolor). U všech rostlin byly měřeny čtyři parametry: délka a šířka okvětních lístků (petal) a okvětních plátků (sepal) v centimetrech.

4. Analyzujte délku okvětních plátků (Sepal.Length) u Iris versicolor.
- Určete 30% kvantil délky okvětního plátku Iris versicolor a slovně interpretujte jeho význam. (3b)
 - Určete bodový a 95% intervalový odhad pravděpodobnosti, že délka okvětního plátku u Iris versicolor bude menší než 6 cm. (4b)
 - Na hladině významnosti 5 % ověřte, zda je střední hodnota (resp. medián) délky okvětního plátku u Iris versicolor menší než 6 cm. (3b)

X ... délka okvětního plátku iris versicolor (cm)
 Data měřena p' odlehla' pozorováním

a) $\hat{x}_{0,3} = 5,600$ cm

cca 30% kosatců druhu iris versicolor má délku okv. plátku menší než 5,600 cm.

b) π ... p-ct, že délka okv. plátku iris vers. bude menší než 6 cm

$$\hat{\pi} = \frac{26}{50} = \underline{\underline{0,520}} \quad (= p)$$

Ověřím předpokladu pro 10:

$$n > \frac{9}{p(1-p)} \quad \dots \quad 50 > \frac{9}{0,52 \cdot 0,48} \Rightarrow 50 > 36 \quad \checkmark$$

95% Clopperův - Pearsonův 10: (0,374; 0,664)
 (binom. test (26,5))

cca 52,0% kosatců typu iris versicolor má délku okv. plátku menší než 6 cm. Dle 95% Clopperova - Pearsonova odhadu očekávám, že danou podmínku splňuje 37,4% až 66,4% těchto kosatců.

c) Overim' normality:

Na hl. významnosti 5% nelze samostatně předpokládat normalitu délky okrov. plátnu iris versicolor (Shapiro-Wilkův test, p -hodnota = 0,467).

$$H_0: \mu = 6,000 \text{ cm}$$

$$H_A: \mu < 6,000 \text{ cm}$$

Na hl. významnosti 5% nepřevládá H_0 ~~ne~~ ~~prospěš~~ H_A (t-test, $x_{\text{obs}} = -0,844$, $df = 49$, p -hodnota = 0,193). Tj. nelze tvrdit, že průměrná délka okrov. plátnu iris versicolor je st. významně menší než 6,000 cm.

Datový soubor iris.xlsx (R. A. Fisher (1936), <http://homel.vsb.cz/~lit40/DATA/iris.xlsx>). „The use of multiple measurements in taxonomic problems“. Annals of Eugenics. 7(2): 179 (188) obsahuje údaje o 50 rostlinách tří odrůd kosatců (Iris setosa, Iris virginica a Iris versicolor). U všech rostlin byly měřeny čtyři parametry: délka a šířka okvětních lístků (petal) a okvětních plátků (sepal) v centimetrech.

5. Na hladině významnosti 5 % rozhodněte, zda je statisticky významný rozdíl mezi průměrnými délkami okvětních plátků (resp. mediány délek okvětních plátků) jednotlivých druhů kosatců. V případě že ano, nalezněte homogenní skupiny druhů kosatců a seřadte je dle délky okvětních plátků sestupně. (10b)

Data obsahují odd. poz. (iris virginica).
Toto bylo z datového zpracování vyloučeno.
(Odd. poz. bylo identifikováno metodou mitování
hradel.)

Nezávislost : Data jsou nezávislá (každá délka
okv. plátku byla měřena u jiné
rostliny).

Normalita :

(*) Druh	Shapiro-Wilkův test (p-hodnota)
setosa	0,460
virginica	0,057
versicolor	0,465

Na kl. v. 5% nezávislým
předpoklad normality (viz. tab. (*))

Shoda rozptylů : Na kl. v. 5% nezávislým
předpoklad o shodě rozptylů
(Bartlettův test, $\chi_{0,05}^2 = 13,8$, $df = 2$,
 p -hodnota = 0,002).

$$H_0: \chi_{0,5}^{\text{setosa}} = \chi_{0,5}^{\text{virginica}} = \chi_{0,5}^{\text{versicolor}}$$

$H_A: \rightarrow H_0$

na hl. výan. 5% pravěb'm H_0

(Kruskalis - Wallisov test, $\chi_{\text{OBS}} = 101$, $df = 2$,
 p -hodnota $\ll 0,001$).

Post-hoc analysis (Dunnov test):

mediciny mēd tū druhy kosatci a st. výan.

↑
dělě obr. pláthi

lis. St. výanamē ujedlēs medicinu dělě
obv. pláthi map' iris virginica, mēchdujē
iris versicolor, st. výanamē nejkratē
dělě obr. pláthi map' iris setosa.