Základy biostatistiky

(MD710P09)
ak. rok 2008/2009

Karel Zvára

karel.zvara@mff.cuni.cz
http://www.karlin.mff.cuni.cz/∼zvara
katedra pravděpodobnosti a matematické statistiky MFF UK

(naposledy upraveno 9. března 2009)

další míry variability

- **rozpětí** $R = x_{\max} - x_{\min}$ [range]
- **kvartilové rozpětí** $\boxed{R_Q = Q_3 - Q_1}$ [interquartile range]
- **variační koeficient** (nesplňuje ani jeden požadavek) slouží k porovnání variability při různých úrovních [coefficient of variation]

$$\boxed{V_x = \frac{s_x}{\bar{x}}}$$

- **entropie** (pro nominální, požadavky nemají smysl, nezávisí na označení hodnot, jen na jejich relativních četnostech)

$$H = -\sum_{j=1}^{m} \frac{n_j}{n} \ln \frac{n_j}{n}$$

další míry variability

- **rozpětí** $R = x_{\max} - x_{\min}$ [range]
- **kvartilové rozpětí** $\boxed{R_Q = Q_3 - Q_1}$ [interquartile range]
- **variační koeficient** (nesplňuje ani jeden požadavek) slouží k porovnání variability při různých úrovních [coefficient of variation]

$$\boxed{V_x = \frac{s_x}{\bar{x}}}$$

- **entropie** (pro nominální, požadavky nemají smysl, nezávisí na označení hodnot, jen na jejich relativních četnostech)

$$H = -\sum_{j=1}^{m} \frac{n_j}{n} \ln \frac{n_j}{n}$$

další míry variability

- **rozpětí** $R = x_{\max} - x_{\min}$ [range]
- **kvartilové rozpětí** $\boxed{R_Q = Q_3 - Q_1}$ [interquartile range]
- **variační koeficient** (nesplňuje ani jeden požadavek) slouží k porovnání variability při různých úrovních [coefficient of variation]

$$\boxed{V_x = \frac{s_x}{\bar{x}}}$$

- **entropie** (pro nominální, požadavky nemají smysl, nezávisí na označení hodnot, jen na jejich relativních četnostech)

$$H = -\sum_{j=1}^{m} \frac{n_j}{n} \ln \frac{n_j}{n}$$

další míry variability

- **rozpětí** $R = x_{\max} - x_{\min}$ [range]
- **kvartilové rozpětí** $\boxed{R_Q = Q_3 - Q_1}$ [interquartile range]
- **variační koeficient** (nesplňuje ani jeden požadavek) slouží k porovnání variability při různých úrovních [coefficient of variation]

$$\boxed{V_x = \frac{s_x}{\bar{x}}}$$

- **entropie** (pro nominální, požadavky nemají smysl, nezávisí na označení hodnot, jen na jejich relativních četnostech)

$$H = -\sum_{j=1}^{m} \frac{n_j}{n} \ln \frac{n_j}{n}$$

příklad ICHS: vztah mužů ke kouření

vzděl.	vztah ke kouření						celk.	H
	nekuřák/bývalý		střední		silný			
zákl.	25	21,4 %	14	12,0 %	78	66,7 %	117	0,854
odb.	83	28,0 %	24	8,1 %	189	63,9 %	296	0,847
stř.	99	33,2 %	24	8,1 %	175	63, %	298	0,882
VŠ	115	48,3 %	17	7,1 %	106	44,5 %	238	0,900

muži se základním vzděláním:

$$H = -\left(\frac{25}{117}\ln\frac{25}{117} + \frac{14}{117}\ln\frac{14}{117} + \frac{78}{117}\ln\frac{78}{117}\right) = 0{,}854123$$

větší vyrovnanost $\Rightarrow$ větší entropie

z-skóry

- z-**skóry** (normovaná veličina)

$$\boxed{z_i = \frac{x_i - \bar{x}}{s_x}, \qquad i = 1, 2, \ldots, n}$$ [(x-mean(x))/sd(x)]

- hodnoty $z_1, z_2, \ldots, z_n$ „ztratily“ informaci o poloze a variabilitě, vždy platí $\bar{z} = 0, \quad s_z = 1$
- přičtení konstanty ani násobení konstantou z-skóry nezmění
- hodnocení vlastností nezávislých na poloze a variabilitě
- pro data: 4, 5, 8, 9, 10, 13, 21 platí $\bar{x} = 10, s_x = 5{,}715$
- proto dostaneme

$$z_1 = \frac{4 - 10}{5{,}715} = -1{,}050, \ldots, z_7 = \frac{21 - 10}{5{,}715} = 1{,}925$$

z-skóry

- *z*-**skóry** (normovaná veličina)

$$\boxed{z_i = \frac{x_i - \bar{x}}{s_x}, \qquad i = 1, 2, \ldots, n}$$ [(x-mean(x))/sd(x)]

- hodnoty $z_1, z_2, \ldots, z_n$ „ztratily" informaci o poloze a variabilitě, vždy platí $\bar{z} = 0, \quad s_z = 1$
- přičtení konstanty ani násobení konstantou *z*-skóry nezmění
- hodnocení vlastností nezávislých na poloze a variabilitě
- pro data: 4, 5, 8, 9, 10, 13, 21 platí $\bar{x} = 10, s_x = 5{,}715$
- proto dostaneme

$$z_1 = \frac{4 - 10}{5{,}715} = -1{,}050, \ldots, z_7 = \frac{21 - 10}{5{,}715} = 1{,}925$$

z-skóry

- z-**skóry** (normovaná veličina)

$$\boxed{z_i = \frac{x_i - \bar{x}}{s_x}, \qquad i = 1, 2, \ldots, n}$$ [(x-mean(x))/sd(x)]

- hodnoty $z_1, z_2, \ldots, z_n$ „ztratily“ informaci o poloze a variabilitě, vždy platí $\bar{z} = 0, \quad s_z = 1$
- přičtení konstanty ani násobení konstantou z-skóry nezmění
- hodnocení vlastností nezávislých na poloze a variabilitě
- pro data: 4, 5, 8, 9, 10, 13, 21 platí $\bar{x} = 10, s_x = 5{,}715$
- proto dostaneme

$$z_1 = \frac{4 - 10}{5{,}715} = -1{,}050, \ldots, z_7 = \frac{21 - 10}{5{,}715} = 1{,}925$$

z-skóry

- z-**skóry** (normovaná veličina)

$$\boxed{z_i = \frac{x_i - \bar{x}}{s_x}, \qquad i = 1, 2, \ldots, n}$$ [(x-mean(x))/sd(x)]

- hodnoty $z_1, z_2, \ldots, z_n$ „ztratily“ informaci o poloze a variabilitě, vždy platí $\bar{z} = 0, \quad s_z = 1$
- přičtení konstanty ani násobení konstantou z-skóry nezmění
- hodnocení vlastností nezávislých na poloze a variabilitě
- pro data: 4, 5, 8, 9, 10, 13, 21 platí $\bar{x} = 10, s_x = 5{,}715$
- proto dostaneme

$$z_1 = \frac{4 - 10}{5{,}715} = -1{,}050, \ldots, z_7 = \frac{21 - 10}{5{,}715} = 1{,}925$$

z-skóry

- z-**skóry** (normovaná veličina)

$$\boxed{z_i = \frac{x_i - \bar{x}}{s_x}, \qquad i = 1, 2, \ldots, n}$$ [(x-mean(x))/sd(x)]

- hodnoty $z_1, z_2, \ldots, z_n$ „ztratily" informaci o poloze a variabilitě, vždy platí $\bar{z} = 0, \quad s_z = 1$
- přičtení konstanty ani násobení konstantou z-skóry nezmění
- hodnocení vlastností nezávislých na poloze a variabilitě
- pro data: 4, 5, 8, 9, 10, 13, 21 platí $\bar{x} = 10, s_x = 5{,}715$
- proto dostaneme

$$z_1 = \frac{4 - 10}{5{,}715} = -1{,}050, \ldots, z_7 = \frac{21 - 10}{5{,}715} = 1{,}925$$

z-skóry

- z-**skóry** (normovaná veličina)

$$\boxed{z_i = \frac{x_i - \bar{x}}{s_x}, \qquad i = 1, 2, \ldots, n}$$ [(x-mean(x))/sd(x)]

- hodnoty $z_1, z_2, \ldots, z_n$ „ztratily“ informaci o poloze a variabilitě, vždy platí $\bar{z} = 0, \quad s_z = 1$
- přičtení konstanty ani násobení konstantou z-skóry nezmění
- hodnocení vlastností nezávislých na poloze a variabilitě
- pro data: 4, 5, 8, 9, 10, 13, 21 platí $\bar{x} = 10, s_x = 5{,}715$
- proto dostaneme

$$z_1 = \frac{4 - 10}{5{,}715} = -1{,}050, \ldots, z_7 = \frac{21 - 10}{5{,}715} = 1{,}925$$

šikmost, špičatost

- **šikmost** (průměr 3. mocnin z-skórů)

$$\boxed{g_1 = \frac{1}{n}\sum_{i=1}^{n} z_i^3 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)^3}$$

[mean(((x-mean(x))/sd(x))^3)]

- **špičatost** (průměr 4. mocnin z-skórů zmenšený o 3)

$$\boxed{g_2 = \frac{1}{n}\sum_{i=1}^{n} z_i^4 - 3 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)^4 - 3}$$

[mean(((x-mean(x))/sd(x))^4)-3]

- g_1, g_2 se používají k posouzení normality
- pro data: 4, 5, 8, 9, 10, 13, 21 dostaneme

$$g_1 = 0{,}771 \qquad g_2 = -0{,}770$$

šikmost, špičatost

- **šikmost** (průměr 3. mocnin *z*-skórů)

$$\boxed{g_1 = \frac{1}{n}\sum_{i=1}^{n} z_i^3 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)^3}$$

[mean(((x-mean(x))/sd(x))^3)]

- **špičatost** (průměr 4. mocnin *z*-skórů zmenšený o 3)

$$\boxed{g_2 = \frac{1}{n}\sum_{i=1}^{n} z_i^4 - 3 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)^4 - 3}$$

[mean(((x-mean(x))/sd(x))^4)-3]

- g_1, g_2 se používají k posouzení normality
- pro data: 4, 5, 8, 9, 10, 13, 21 dostaneme

$$g_1 = 0{,}771 \qquad g_2 = -0{,}770$$

šikmost, špičatost

- **šikmost** (průměr 3. mocnin z-skórů)

$$\boxed{g_1 = \frac{1}{n}\sum_{i=1}^{n} z_i^3 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)^3}$$

[mean(((x-mean(x))/sd(x))^3)]

- **špičatost** (průměr 4. mocnin z-skórů zmenšený o 3)

$$\boxed{g_2 = \frac{1}{n}\sum_{i=1}^{n} z_i^4 - 3 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)^4 - 3}$$

[mean(((x-mean(x))/sd(x))^4)-3]

- g_1, g_2 se používají k posouzení normality
- pro data: 4, 5, 8, 9, 10, 13, 21 dostaneme

$$g_1 = 0{,}771 \qquad g_2 = -0{,}770$$

šikmost, špičatost

- **šikmost** (průměr 3. mocnin z-skórů)

$$\boxed{g_1 = \frac{1}{n}\sum_{i=1}^{n} z_i^3 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)^3}$$

[mean(((x-mean(x))/sd(x))^3)]

- **špičatost** (průměr 4. mocnin z-skórů zmenšený o 3)

$$\boxed{g_2 = \frac{1}{n}\sum_{i=1}^{n} z_i^4 - 3 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)^4 - 3}$$

[mean(((x-mean(x))/sd(x))^4)-3]

- g_1, g_2 se používají k posouzení normality
- pro data: 4, 5, 8, 9, 10, 13, 21 dostaneme

$$g_1 = 0{,}771 \qquad g_2 = -0{,}770$$

normální diagram

[(normal) probability plot], [quantile-comparison plot]

- k ověřování předpokladu **normálního** rozdělení
- porovnává skutečnou variační řadu s ideální řadou normálního (Gaussova) rozdělení
- v ideálním případě body téměř na přímce
- systematická odchylka ukazuje na rozdělení, které není normální
- konvexní či konkávní průběh – nesymetrie (nenulová šikmost)
- esovitý průběh – nenulová špičatost
- [qqnorm(x)]
- přímku vloží [qqline(x)]

normální diagram

[(normal) probability plot], [quantile-comparison plot]

- k ověřování předpokladu **normálního** rozdělení
- porovnává skutečnou variační řadu s ideální řadou normálního (Gaussova) rozdělení
- v ideálním případě body téměř na přímce
- systematická odchylka ukazuje na rozdělení, které není normální
- konvexní či konkávní průběh – nesymetrie (nenulová šikmost)
- esovitý průběh – nenulová špičatost
- [qqnorm(x)]
- přímku vloží [qqline(x)]

normální diagram

[(normal) probability plot], [quantile-comparison plot]

- k ověřování předpokladu **normálního** rozdělení
- porovnává skutečnou variační řadu s ideální řadou normálního (Gaussova) rozdělení
- v ideálním případě body téměř na přímce
- systematická odchylka ukazuje na rozdělení, které není normální
- konvexní či konkávní průběh – nesymetrie (nenulová šikmost)
- esovitý průběh – nenulová špičatost
- [qqnorm(x)]
- přímku vloží [qqline(x)]

normální diagram

[(normal) probability plot], [quantile-comparison plot]

- k ověřování předpokladu **normálního** rozdělení
- porovnává skutečnou variační řadu s ideální řadou normálního (Gaussova) rozdělení
- v ideálním případě body téměř na přímce
- systematická odchylka ukazuje na rozdělení, které není normální
- konvexní či konkávní průběh – nesymetrie (nenulová šikmost)
- esovitý průběh – nenulová špičatost
- [qqnorm(x)]
- přímku vloží [qqline(x)]

normální diagram

[(normal) probability plot], [quantile-comparison plot]

- k ověřování předpokladu **normálního** rozdělení
- porovnává skutečnou variační řadu s ideální řadou normálního (Gaussova) rozdělení
- v ideálním případě body téměř na přímce
- systematická odchylka ukazuje na rozdělení, které není normální
- konvexní či konkávní průběh – nesymetrie (nenulová šikmost)
- esovitý průběh – nenulová špičatost
- [qqnorm(x)]
- přímku vloží [qqline(x)]

normální diagram

[(normal) probability plot], [quantile-comparison plot]

- k ověřování předpokladu **normálního** rozdělení
- porovnává skutečnou variační řadu s ideální řadou normálního (Gaussova) rozdělení
- v ideálním případě body téměř na přímce
- systematická odchylka ukazuje na rozdělení, které není normální
- konvexní či konkávní průběh – nesymetrie (nenulová šikmost)
- esovitý průběh – nenulová špičatost
- [qqnorm(x)]
- přímku vloží [qqline(x)]

normální diagram

[(normal) probability plot], [quantile-comparison plot]

- k ověřování předpokladu **normálního** rozdělení
- porovnává skutečnou variační řadu s ideální řadou normálního (Gaussova) rozdělení
- v ideálním případě body téměř na přímce
- systematická odchylka ukazuje na rozdělení, které není normální
- konvexní či konkávní průběh – nesymetrie (nenulová šikmost)
- esovitý průběh – nenulová špičatost
- [qqnorm(x)]
- přímku vloží [qqline(x)]

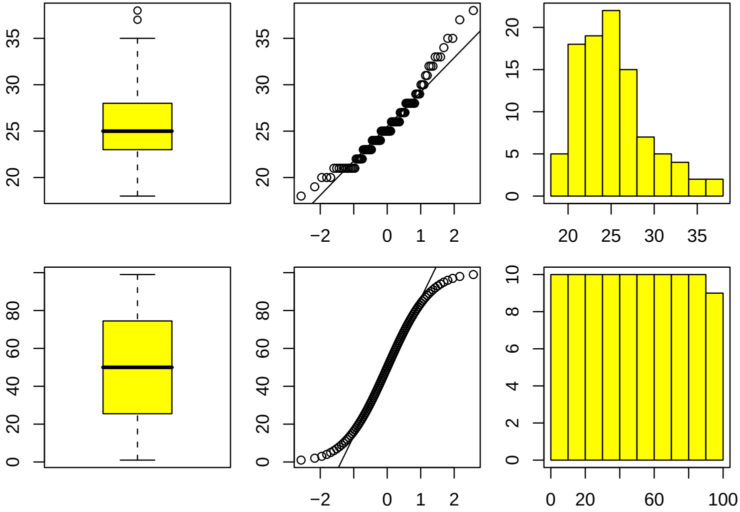
normální diagram

[(normal) probability plot], [quantile-comparison plot]

- k ověřování předpokladu **normálního** rozdělení
- porovnává skutečnou variační řadu s ideální řadou normálního (Gaussova) rozdělení
- v ideálním případě body téměř na přímce
- systematická odchylka ukazuje na rozdělení, které není normální
- konvexní či konkávní průběh – nesymetrie (nenulová šikmost)
- esovitý průběh – nenulová špičatost
- [qqnorm(x)]
- přímku vloží [qqline(x)]

příklad: věk matky, čísla 1 až 99

věk matek: $g_1 = 0{,}741, \quad g_2 = 0{,}220$ čísla 1 až 99: $g_1 = 0, \quad g_2 = -1{,}236$



závislost dvojice znaků

- možnost zkoumání závislosti dvou znaků
- způsob znázornění (prokazování) závisí na měřítcích znaků
- **kvantitativní – kvantitativní**
  rozptylový (bodový) diagram [scatter plot]
  korelace, regrese [correlation, regression]
- **kvantitativní - kvalitativní**
  krabicový diagram [box-plot]
  t-test, ANOVA
- **kvalitativní - kvalitativní**
  kontingenční tabulka [contingency table]
  chí-kvadrát test, Fisherův exaktní test

závislost dvojice znaků

- možnost zkoumání závislosti dvou znaků
- způsob znázornění (prokazování) závisí na měřítcích znaků
- **kvantitativní – kvantitativní**
  rozptylový (bodový) diagram [scatter plot]
  korelace, regrese [correlation, regression]
- **kvantitativní - kvalitativní**
  krabicový diagram [box-plot]
  t-test, ANOVA
- **kvalitativní - kvalitativní**
  kontingenční tabulka [contingency table]
  chí-kvadrát test, Fisherův exaktní test

závislost dvojice znaků

- možnost zkoumání závislosti dvou znaků
- způsob znázornění (prokazování) závisí na měřítcích znaků
- **kvantitativní – kvantitativní**
  rozptylový (bodový) diagram [scatter plot]
  korelace, regrese [correlation, regression]
- **kvantitativní - kvalitativní**
  krabicový diagram [box-plot]
  t-test, ANOVA
- **kvalitativní - kvalitativní**
  kontingenční tabulka [contingency table]
  chí-kvadrát test, Fisherův exaktní test

závislost dvojice znaků

- možnost zkoumání závislosti dvou znaků
- způsob znázornění (prokazování) závisí na měřítcích znaků
- **kvantitativní – kvantitativní**
  rozptylový (bodový) diagram [scatter plot]
  korelace, regrese [correlation, regression]
- **kvantitativní - kvalitativní**
  krabicový diagram [box-plot]
  t-test, ANOVA
- **kvalitativní - kvalitativní**
  kontingenční tabulka [contingency table]
  chí-kvadrát test, Fisherův exaktní test

závislost dvojice znaků

- možnost zkoumání závislosti dvou znaků
- způsob znázornění (prokazování) závisí na měřítcích znaků
- **kvantitativní – kvantitativní**
  rozptylový (bodový) diagram [scatter plot]
  korelace, regrese [correlation, regression]
- **kvantitativní - kvalitativní**
  krabicový diagram [box-plot]
  t-test, ANOVA
- **kvalitativní - kvalitativní**
  kontingenční tabulka [contingency table]
  chí-kvadrát test, Fisherův exaktní test

kvantitativní – kvantitativní

- pokud záleží na směru závislosti, pak vysvětlovanou (**závisle proměnnou**) veličinu umístíme na svislou osu y
- **korelační koeficient** vyjadřuje sílu a směr **vzájemné** závislosti

$$\boxed{r_{xy} = \frac{s_{xy}}{s_x \cdot s_y}}, \quad \text{kde} \quad s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

- [cor(x,y)] [correlation coefficient]
- s_{xy} – výběrová **kovariance** [covariance]
- pomocí z-skórů (nezávislost na poloze a měřítku)

$$\boxed{r_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)}$$

- pro $r_{xy} > 0$ s rostoucím x v průměru roste y
  pro $r_{xy} < 0$ s rostoucím x v průměru klesá y

$$\boxed{-1 \leq r_{xy} \leq 1}$$

kvantitativní – kvantitativní

- pokud záleží na směru závislosti, pak vysvětlovanou (**závisle proměnnou**) veličinu umístíme na svislou osu y
- **korelační koeficient** vyjadřuje sílu a směr **vzájemné** závislosti

$$\boxed{r_{xy} = \frac{s_{xy}}{s_x \cdot s_y}}, \quad \text{kde} \quad s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

- [cor(x,y)] [correlation coefficient]
- s_{xy} – výběrová **kovariance** [covariance]
- pomocí z-skórů (nezávislost na poloze a měřítku)

$$\boxed{r_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)}$$

- pro $r_{xy} > 0$ s rostoucím x v průměru roste y
  pro $r_{xy} < 0$ s rostoucím x v průměru klesá y

$$\boxed{-1 \leq r_{xy} \leq 1}$$

kvantitativní – kvantitativní

- pokud záleží na směru závislosti, pak vysvětlovanou (**závisle proměnnou**) veličinu umístíme na svislou osu y
- **korelační koeficient** vyjadřuje sílu a směr **vzájemné** závislosti

$$\boxed{r_{xy} = \frac{s_{xy}}{s_x \cdot s_y}}, \quad \text{kde} \quad s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

- [cor(x,y)] [correlation coefficient]
- s_{xy} – výběrová **kovariance** [covariance]
- pomocí z-skórů (nezávislost na poloze a měřítku)

$$\boxed{r_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)}$$

- pro $r_{xy} > 0$ s rostoucím x v průměru roste y
  pro $r_{xy} < 0$ s rostoucím x v průměru klesá y

$$\boxed{-1 \le r_{xy} \le 1}$$

kvantitativní – kvantitativní

- pokud záleží na směru závislosti, pak vysvětlovanou (**závisle proměnnou**) veličinu umístíme na svislou osu y
- **korelační koeficient** vyjadřuje sílu a směr **vzájemné** závislosti

$$\boxed{r_{xy} = \frac{s_{xy}}{s_x \cdot s_y}}, \quad \text{kde} \quad s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

- [cor(x,y)] [correlation coefficient]
- s_{xy} – výběrová **kovariance** [covariance]
- pomocí z-skórů (nezávislost na poloze a měřítku)

$$\boxed{r_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)}$$

- pro $r_{xy} > 0$ s rostoucím x v průměru roste y
  pro $r_{xy} < 0$ s rostoucím x v průměru klesá y

$$\boxed{-1 \leq r_{xy} \leq 1}$$

kvantitativní – kvantitativní

- pokud záleží na směru závislosti, pak vysvětlovanou (**závisle proměnnou**) veličinu umístíme na svislou osu y
- **korelační koeficient** vyjadřuje sílu a směr **vzájemné** závislosti

$$\boxed{r_{xy} = \frac{s_{xy}}{s_x \cdot s_y}}, \quad \text{kde} \quad s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

- [cor(x,y)] [correlation coefficient]
- s_{xy} – výběrová **kovariance** [covariance]
- pomocí z-skórů (nezávislost na poloze a měřítku)

$$\boxed{r_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)}$$

- pro $r_{xy} > 0$ s rostoucím x v průměru roste y
  pro $r_{xy} < 0$ s rostoucím x v průměru klesá y

$$\boxed{-1 \leq r_{xy} \leq 1}$$

kvantitativní – kvantitativní

- pokud záleží na směru závislosti, pak vysvětlovanou (**závisle proměnnou**) veličinu umístíme na svislou osu y
- **korelační koeficient** vyjadřuje sílu a směr **vzájemné** závislosti

$$\boxed{r_{xy} = \frac{s_{xy}}{s_x \cdot s_y}}, \quad \text{kde} \quad s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

- [cor(x,y)] [correlation coefficient]
- s_{xy} – výběrová **kovariance** [covariance]
- pomocí z-skórů (nezávislost na poloze a měřítku)

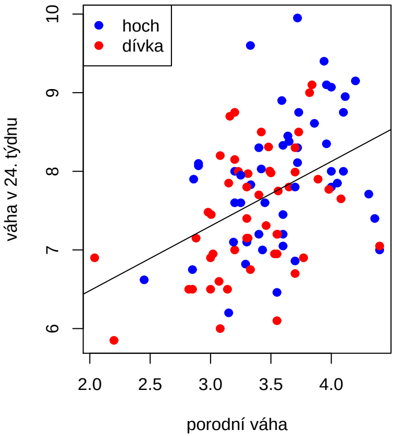
$$\boxed{r_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)}$$

- pro $r_{xy} > 0$ s rostoucím x v průměru roste y
  pro $r_{xy} < 0$ s rostoucím x v průměru klesá y

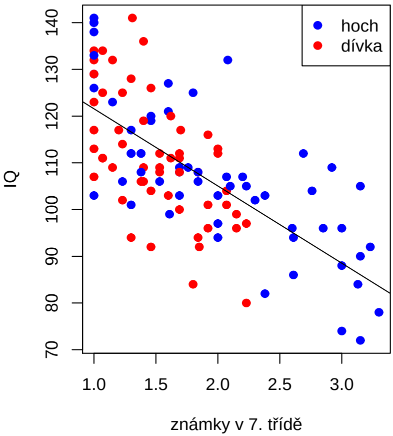
$$\boxed{-1 \leq r_{xy} \leq 1}$$

kvantitativní – kvantitativní, příklady

vlevo – závislost váhy v 24. týdnu na porodní váze s rozlišením pohlaví (data: Kojení)
vpravo – závislost IQ na průměrné známce v 7. třídě (data: Iq3)



$r = 0{,}429$



$r = -0{,}689$

kvalitativní – kvalitativní

- **kontingenční tabulka** [contingency table]
  obsahuje přehledně zapsané úplné údaje
- sdružené četnosti jednotlivých kombinací hodnot dvou znaků
- marginální četnosti:
  - **řádkové** marginální četnosti: součty sdružených četností v jednotlivých řádcích (pro jednotlivé hodnoty řádkového znaku)
  - **sloupcové** marginální četnosti: součty sdružených četností v jednotlivých sloupcích (pro jednotlivé hodnoty sloupcového znaku)
- [table(F,G)] nebo [xtabs(∼ F + G)]
  resp. [xtabs(∼ F + G , data=DataFrame)]
  kde F a G jsou v R faktory, DataFrame je databáze

kvalitativní – kvalitativní

- **kontingenční tabulka** [contingency table]
  obsahuje přehledně zapsané úplné údaje
- sdružené četnosti jednotlivých kombinací hodnot dvou znaků
- marginální četnosti:
  - **řádkové** marginální četnosti: součty sdružených četností v jednotlivých řádcích (pro jednotlivé hodnoty řádkového znaku)
  - **sloupcové** marginální četnosti: součty sdružených četností v jednotlivých sloupcích (pro jednotlivé hodnoty sloupcového znaku)
- [table(F,G)] nebo [xtabs(∼ F + G)]
  resp. [xtabs(∼ F + G , data=DataFrame)]
  kde F a G jsou v R faktory, DataFrame je databáze

kvalitativní – kvalitativní

- **kontingenční tabulka** [contingency table]
  obsahuje přehledně zapsané úplné údaje
- sdružené četnosti jednotlivých kombinací hodnot dvou znaků
- marginální četnosti:
  - **řádkové** marginální četnosti: součty sdružených četností v jednotlivých řádcích (pro jednotlivé hodnoty řádkového znaku)
  - **sloupcové** marginální četnosti: součty sdružených četností v jednotlivých sloupcích (pro jednotlivé hodnoty sloupcového znaku)
- [table(F,G)] nebo [xtabs(∼ F + G)]
  resp. [xtabs(∼ F + G , data=DataFrame)]
  kde F a G jsou v R faktory, DataFrame je databáze

kvalitativní – kvalitativní

- **kontingenční tabulka** [contingency table]
  obsahuje přehledně zapsané úplné údaje
- sdružené četnosti jednotlivých kombinací hodnot dvou znaků
- marginální četnosti:
  - **řádkové** marginální četnosti: součty sdružených četností v jednotlivých řádcích (pro jednotlivé hodnoty řádkového znaku)
  - **sloupcové** marginální četnosti: součty sdružených četností v jednotlivých sloupcích (pro jednotlivé hodnoty sloupcového znaku)
- [table(F,G)] nebo [xtabs(∼ F + G)]
  resp. [xtabs(∼ F + G , data=DataFrame)]
  kde F a G jsou v R faktory, DataFrame je databáze

kvalitativní – kvalitativní

- **kontingenční tabulka** [contingency table]
  obsahuje přehledně zapsané úplné údaje
- sdružené četnosti jednotlivých kombinací hodnot dvou znaků
- marginální četnosti:
  - **řádkové** marginální četnosti: součty sdružených četností v jednotlivých řádcích (pro jednotlivé hodnoty řádkového znaku)
  - **sloupcové** marginální četnosti: součty sdružených četností v jednotlivých sloupcích (pro jednotlivé hodnoty sloupcového znaku)
- [table(F,G)] nebo [xtabs(∼ F + G)]
  resp. [xtabs(∼ F + G , data=DataFrame)]
  kde F a G jsou v R faktory, DataFrame je databáze

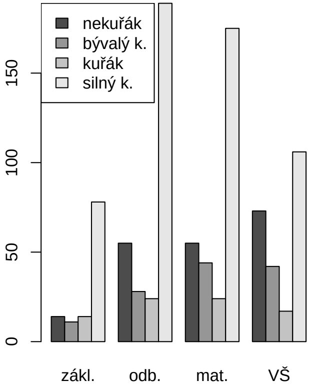
kvalitativní – kvalitativní

- **kontingenční tabulka** [contingency table]
  obsahuje přehledně zapsané úplné údaje
- sdružené četnosti jednotlivých kombinací hodnot dvou znaků
- marginální četnosti:
  - **řádkové** marginální četnosti: součty sdružených četností v jednotlivých řádcích (pro jednotlivé hodnoty řádkového znaku)
  - **sloupcové** marginální četnosti: součty sdružených četností v jednotlivých sloupcích (pro jednotlivé hodnoty sloupcového znaku)
- [table(F,G)] nebo [xtabs(∼ F + G)]
  resp. [xtabs(∼ F + G , data=DataFrame)]
  kde F a G jsou v R faktory, DataFrame je databáze

příklad: kouření u mužů
data: lchs

vzdělání	zákl.	odb.	mat.	VŠ	celk.
nekuřák	14	55	55	73	197
bývalý k.	11	28	44	42	125
kuřák	14	24	24	17	79
silný k.	78	189	175	106	548
celkem	117	296	298	238	949

v grafu znázorněny **absolutní** četnosti
(sdružené, marginální četnosti)
[barplot(t,beside=TRUE)]



relativní četnosti v kontingenční tabulce

- **řádková procenta** (relativní četnosti v daném řádku)
  - podíl jednotlivých hodnot sloupcového znaku pro danou hodnotu řádkového znaku
  - **podmíněné rozdělení** hodnot sloupcového znaku pro danou hodnotu řádkového znaku
- **sloupcová procenta** (relativní četnosti v daném sloupci)
  - podíl jednotlivých hodnot řádkového znaku pro danou hodnotu sloupcového znaku
  - **podmíněné rozdělení** hodnot řádkového znaku pro danou hodnotu sloupcového znaku
- **nezávislosti** obou znaků odpovídá situace, kdy jsou např. sloupcová procenta pro všechny hodnoty sloupcového znaku podobné

relativní četnosti v kontingenční tabulce

- **řádková procenta** (relativní četnosti v daném řádku)
  - podíl jednotlivých hodnot sloupcového znaku pro danou hodnotu řádkového znaku
  - **podmíněné rozdělení** hodnot sloupcového znaku pro danou hodnotu řádkového znaku
- **sloupcová procenta** (relativní četnosti v daném sloupci)
  - podíl jednotlivých hodnot řádkového znaku pro danou hodnotu sloupcového znaku
  - **podmíněné rozdělení** hodnot řádkového znaku pro danou hodnotu sloupcového znaku
- **nezávislosti** obou znaků odpovídá situace, kdy jsou např. sloupcová procenta pro všechny hodnoty sloupcového znaku podobné

relativní četnosti v kontingenční tabulce

- **řádková procenta** (relativní četnosti v daném řádku)
  - podíl jednotlivých hodnot sloupcového znaku pro danou hodnotu řádkového znaku
  - **podmíněné rozdělení** hodnot sloupcového znaku pro danou hodnotu řádkového znaku
- **sloupcová procenta** (relativní četnosti v daném sloupci)
  - podíl jednotlivých hodnot řádkového znaku pro danou hodnotu sloupcového znaku
  - **podmíněné rozdělení** hodnot řádkového znaku pro danou hodnotu sloupcového znaku
- **nezávislosti** obou znaků odpovídá situace, kdy jsou např. sloupcová procenta pro všechny hodnoty sloupcového znaku podobné

relativní četnosti v kontingenční tabulce

- **řádková procenta** (relativní četnosti v daném řádku)
  - podíl jednotlivých hodnot sloupcového znaku pro danou hodnotu řádkového znaku
  - **podmíněné rozdělení** hodnot sloupcového znaku pro danou hodnotu řádkového znaku
- **sloupcová procenta** (relativní četnosti v daném sloupci)
  - podíl jednotlivých hodnot řádkového znaku pro danou hodnotu sloupcového znaku
  - **podmíněné rozdělení** hodnot řádkového znaku pro danou hodnotu sloupcového znaku
- **nezávislosti** obou znaků odpovídá situace, kdy jsou např. sloupcová procenta pro všechny hodnoty sloupcového znaku podobné

relativní četnosti v kontingenční tabulce

- **řádková procenta** (relativní četnosti v daném řádku)
  - podíl jednotlivých hodnot sloupcového znaku pro danou hodnotu řádkového znaku
  - **podmíněné rozdělení** hodnot sloupcového znaku pro danou hodnotu řádkového znaku
- **sloupcová procenta** (relativní četnosti v daném sloupci)
  - podíl jednotlivých hodnot řádkového znaku pro danou hodnotu sloupcového znaku
  - **podmíněné rozdělení** hodnot řádkového znaku pro danou hodnotu sloupcového znaku
- **nezávislosti** obou znaků odpovídá situace, kdy jsou např. sloupcová procenta pro všechny hodnoty sloupcového znaku podobné

relativní četnosti v kontingenční tabulce

- **řádková procenta** (relativní četnosti v daném řádku)
  - podíl jednotlivých hodnot sloupcového znaku pro danou hodnotu řádkového znaku
  - **podmíněné rozdělení** hodnot sloupcového znaku pro danou hodnotu řádkového znaku
- **sloupcová procenta** (relativní četnosti v daném sloupci)
  - podíl jednotlivých hodnot řádkového znaku pro danou hodnotu sloupcového znaku
  - **podmíněné rozdělení** hodnot řádkového znaku pro danou hodnotu sloupcového znaku
- **nezávislosti** obou znaků odpovídá situace, kdy jsou např. sloupcová procenta pro všechny hodnoty sloupcového znaku podobné

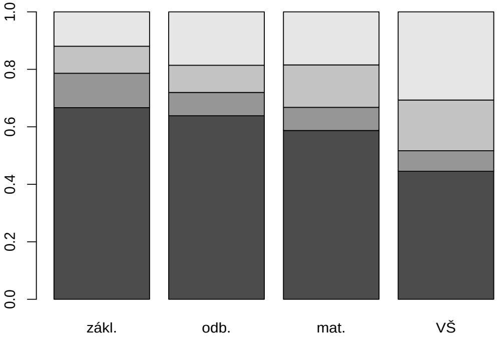
relativní četnosti v kontingenční tabulce

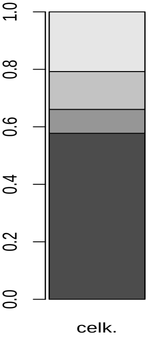
- **řádková procenta** (relativní četnosti v daném řádku)
  - podíl jednotlivých hodnot sloupcového znaku pro danou hodnotu řádkového znaku
  - **podmíněné rozdělení** hodnot sloupcového znaku pro danou hodnotu řádkového znaku
- **sloupcová procenta** (relativní četnosti v daném sloupci)
  - podíl jednotlivých hodnot řádkového znaku pro danou hodnotu sloupcového znaku
  - **podmíněné rozdělení** hodnot řádkového znaku pro danou hodnotu sloupcového znaku
- **nezávislosti** obou znaků odpovídá situace, kdy jsou např. sloupcová procenta pro všechny hodnoty sloupcového znaku podobné

příklad: kouření u mužů

podmíněné relativní četnosti marginální relativní četnosti

vzdělání	zákl.	odb.	mat.	VŠ	celk.
nekuřák	12,0 %	18,6 %	18,5 %	30,7 %	20,6 %
bývalý k.	9,4 %	9,5 %	14,8 %	17,6 %	13,2 %
kuřák	12,0 %	8,1 %	8,1 %	7,1 %	8,3 %
silný k.	66,7 %	63,9 %	58,7 %	44,5 %	57,8%
celkem	100%	100%	100%	100%	100%

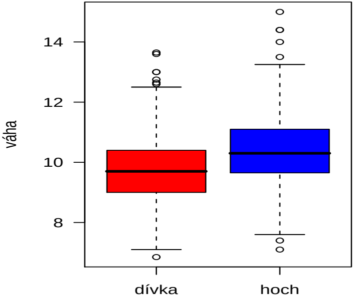


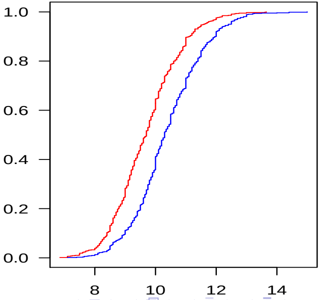


kvantitativní – kvalitativní

váha v jednom roce podle pohlaví, data: Deti1633

- lze chápat jako závislost spojité veličiny na kvalitativní
- srovnání souborů dat (spojitá veličina)
- krabicové diagramy resp. empirické distribuční funkce
- příklad: hmotnost chlapců a dívek v jednom roce
- **nezávislosti** odpovídá podobné umístění krabic resp. empirických distribučních funkcí

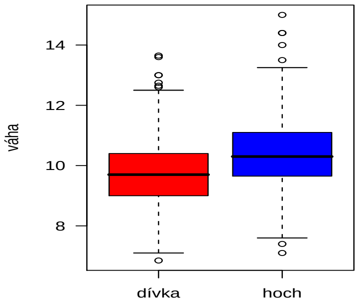


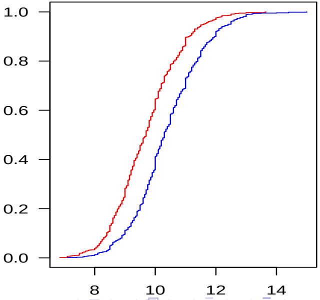


kvantitativní – kvalitativní

váha v jednom roce podle pohlaví, data: Deti1633

- lze chápat jako závislost spojité veličiny na kvalitativní
- srovnání souborů dat (spojitá veličina)
- krabicové diagramy resp. empirické distribuční funkce
- příklad: hmotnost chlapců a dívek v jednom roce
- **nezávislosti** odpovídá podobné umístění krabic resp. empirických distribučních funkcí

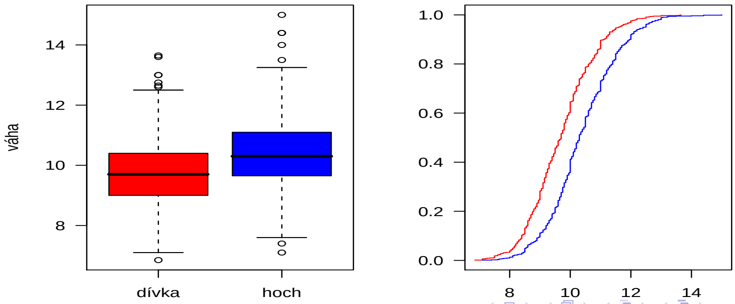




kvantitativní – kvalitativní

váha v jednom roce podle pohlaví, data: Deti1633

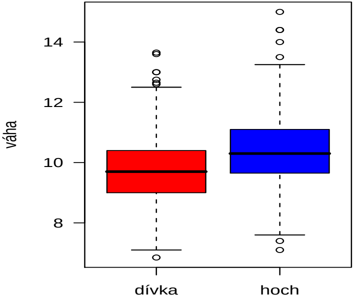
- lze chápat jako závislost spojité veličiny na kvalitativní
- srovnání souborů dat (spojitá veličina)
- krabicové diagramy resp. empirické distribuční funkce
- příklad: hmotnost chlapců a dívek v jednom roce
- **nezávislosti** odpovídá podobné umístění krabic resp. empirických distribučních funkcí

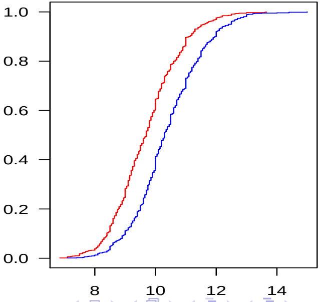


kvantitativní – kvalitativní

váha v jednom roce podle pohlaví, data: Deti1633

- lze chápat jako závislost spojité veličiny na kvalitativní
- srovnání souborů dat (spojitá veličina)
- krabicové diagramy resp. empirické distribuční funkce
- příklad: hmotnost chlapců a dívek v jednom roce
- **nezávislosti** odpovídá podobné umístění krabic resp. empirických distribučních funkcí

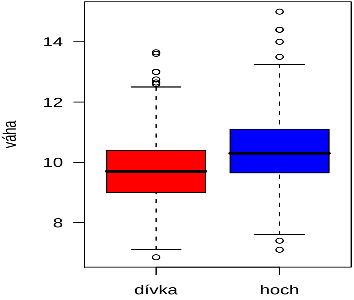


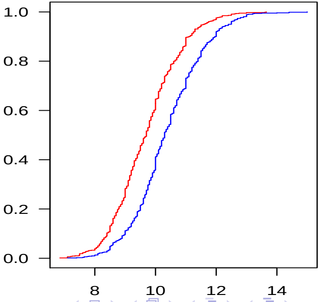


kvantitativní – kvalitativní

váha v jednom roce podle pohlaví, data: Deti1633

- lze chápat jako závislost spojité veličiny na kvalitativní
- srovnání souborů dat (spojitá veličina)
- krabicové diagramy resp. empirické distribuční funkce
- příklad: hmotnost chlapců a dívek v jednom roce
- **nezávislosti** odpovídá podobné umístění krabic resp. empirických distribučních funkcí

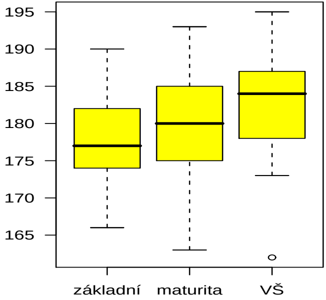


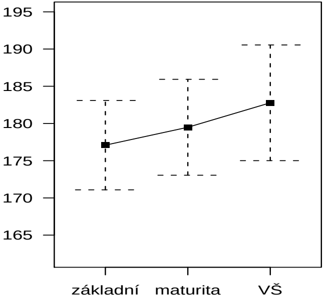


příklad: závislost výšky otce na vzdělání matky

data: Kojení

- porovnáme výšky otců ve skupinách podle vzdělání matky
- napravo znázorníme průměry a směrodatné odchylky
- intervaly kolem průměru mívají i jinou interpretaci (jsou jiné)

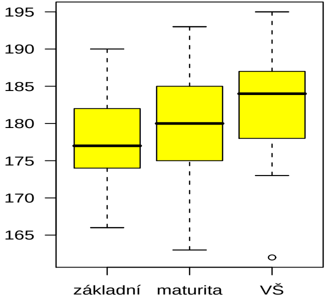


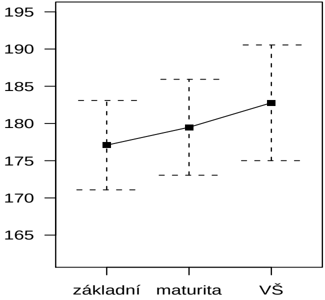


příklad: závislost výšky otce na vzdělání matky

data: Kojení

- porovnáme výšky otců ve skupinách podle vzdělání matky
- napravo znázorníme průměry a směrodatné odchylky
- intervaly kolem průměru mívají i jinou interpretaci (jsou jiné)

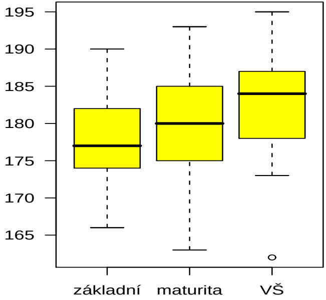


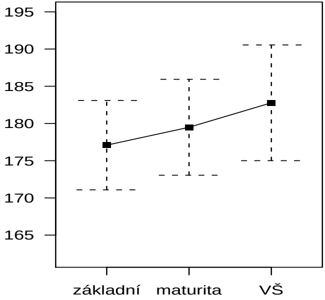


příklad: závislost výšky otce na vzdělání matky

data: Kojení

- porovnáme výšky otců ve skupinách podle vzdělání matky
- napravo znázorníme průměry a směrodatné odchylky
- intervaly kolem průměru mívají i jinou interpretaci (jsou jiné)





Náhodné jevy

- **náhodný pokus** výsledek předem neurčitý
- předpokládá se **stabilita relativních četností** možných výsledků, která s nezávislými opakováními pokusu roste
- **náhodný jev** tvrzení o výsledku náhodného pokusu (podmnožina množiny Ω)
- **jistý jev** Ω nastává vždy
- **nemožný jev** $\emptyset$ nenastává nikdy
- **podjev**: $B \subset D$ znamená $B \Rightarrow D$
- **jev opačný**: $\overline{D} \Leftrightarrow$ neplatí D
- **průnik jevů** $B \cap D$ nastaly oba jevy
- **sjednocení jevů** $D \cup B$ nastal **aspoň** jeden
- **neslučitelné jevy** $B \cap D = \emptyset$

Náhodné jevy

- **náhodný pokus** výsledek předem neurčitý
- předpokládá se **stabilita relativních četností** možných výsledků, která s nezávislými opakováními pokusu roste
- **náhodný jev** tvrzení o výsledku náhodného pokusu (podmnožina množiny Ω)
- **jistý jev** Ω nastává vždy
- **nemožný jev** $\emptyset$ nenastává nikdy
- **podjev**: $B \subset D$ znamená $B \Rightarrow D$
- **jev opačný**: $\overline{D} \Leftrightarrow$ neplatí D
- **průnik jevů** $B \cap D$ nastaly oba jevy
- **sjednocení jevů** $D \cup B$ nastal **aspoň** jeden
- **neslučitelné jevy** $B \cap D = \emptyset$

Náhodné jevy

- **náhodný pokus** výsledek předem neurčitý
- předpokládá se **stabilita relativních četností** možných výsledků, která s nezávislými opakováními pokusu roste
- **náhodný jev** tvrzení o výsledku náhodného pokusu (podmnožina množiny Ω)
- **jistý jev** Ω nastává vždy
- **nemožný jev** $\emptyset$ nenastává nikdy
- **podjev**: $B \subset D$ znamená $B \Rightarrow D$
- **jev opačný**: $\overline{D} \Leftrightarrow$ neplatí D
- **průnik jevů** $B \cap D$ nastaly oba jevy
- **sjednocení jevů** $D \cup B$ nastal **aspoň** jeden
- **neslučitelné jevy** $B \cap D = \emptyset$

Náhodné jevy

- **náhodný pokus** výsledek předem neurčitý
- předpokládá se **stabilita relativních četností** možných výsledků, která s nezávislými opakováními pokusu roste
- **náhodný jev** tvrzení o výsledku náhodného pokusu (podmnožina množiny Ω)
- **jistý jev** Ω nastává vždy
- **nemožný jev** $\emptyset$ nenastává nikdy
- **podjev**: $B \subset D$ znamená $B \Rightarrow D$
- **jev opačný**: $\overline{D} \Leftrightarrow$ neplatí D
- **průnik jevů** $B \cap D$ nastaly oba jevy
- **sjednocení jevů** $D \cup B$ nastal **aspoň** jeden
- **neslučitelné jevy** $B \cap D = \emptyset$

Náhodné jevy

- **náhodný pokus** výsledek předem neurčitý
- předpokládá se **stabilita relativních četností** možných výsledků, která s nezávislými opakováními pokusu roste
- **náhodný jev** tvrzení o výsledku náhodného pokusu (podmnožina množiny Ω)
- **jistý jev** Ω nastává vždy
- **nemožný jev** $\emptyset$ nenastává nikdy
- **podjev**: $B \subset D$ znamená $B \Rightarrow D$
- **jev opačný**: $\overline{D} \Leftrightarrow$ neplatí D
- **průnik jevů** $B \cap D$ nastaly oba jevy
- **sjednocení jevů** $D \cup B$ nastal **aspoň** jeden
- **neslučitelné jevy** $B \cap D = \emptyset$

Náhodné jevy

- **náhodný pokus** výsledek předem neurčitý
- předpokládá se **stabilita relativních četností** možných výsledků, která s nezávislými opakováními pokusu roste
- **náhodný jev** tvrzení o výsledku náhodného pokusu (podmnožina množiny Ω)
- **jistý jev** Ω nastává vždy
- **nemožný jev** $\emptyset$ nenastává nikdy
- **podjev**: $B \subset D$ znamená $B \Rightarrow D$
- **jev opačný**: $\overline{D} \Leftrightarrow$ neplatí D
- **průnik jevů** $B \cap D$ nastaly oba jevy
- **sjednocení jevů** $D \cup B$ nastal **aspoň** jeden
- **neslučitelné jevy** $B \cap D = \emptyset$

Náhodné jevy

- **náhodný pokus** výsledek předem neurčitý
- předpokládá se **stabilita relativních četností** možných výsledků, která s nezávislými opakováními pokusu roste
- **náhodný jev** tvrzení o výsledku náhodného pokusu (podmnožina množiny Ω)
- **jistý jev** Ω nastává vždy
- **nemožný jev** $\emptyset$ nenastává nikdy
- **podjev**: $B \subset D$ znamená $B \Rightarrow D$
- **jev opačný**: $\overline{D} \Leftrightarrow$ neplatí D
- **průnik jevů** $B \cap D$ nastaly oba jevy
- **sjednocení jevů** $D \cup B$ nastal **aspoň** jeden
- **neslučitelné jevy** $B \cap D = \emptyset$

Náhodné jevy

- **náhodný pokus** výsledek předem neurčitý
- předpokládá se **stabilita relativních četností** možných výsledků, která s nezávislými opakováními pokusu roste
- **náhodný jev** tvrzení o výsledku náhodného pokusu (podmnožina množiny Ω)
- **jistý jev** Ω nastává vždy
- **nemožný jev** $\emptyset$ nenastává nikdy
- **podjev**: $B \subset D$ znamená $B \Rightarrow D$
- **jev opačný**: $\overline{D} \Leftrightarrow$ neplatí D
- **průnik jevů** $B \cap D$ nastaly oba jevy
- **sjednocení jevů** $D \cup B$ nastal **aspoň** jeden
- **neslučitelné jevy** $B \cap D = \emptyset$

Náhodné jevy

- **náhodný pokus** výsledek předem neurčitý
- předpokládá se **stabilita relativních četností** možných výsledků, která s nezávislými opakováními pokusu roste
- **náhodný jev** tvrzení o výsledku náhodného pokusu (podmnožina množiny Ω)
- **jistý jev** Ω nastává vždy
- **nemožný jev** $\emptyset$ nenastává nikdy
- **podjev**: $B \subset D$ znamená $B \Rightarrow D$
- **jev opačný**: $\overline{D} \Leftrightarrow$ neplatí D
- **průnik jevů** $B \cap D$ nastaly oba jevy
- **sjednocení jevů** $D \cup B$ nastal **aspoň** jeden
- **neslučitelné jevy** $B \cap D = \emptyset$

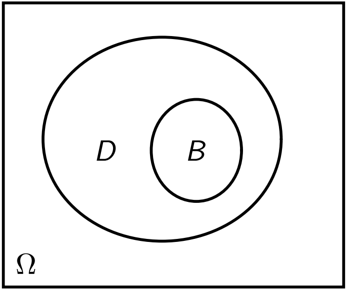
Náhodné jevy

- **náhodný pokus** výsledek předem neurčitý
- předpokládá se **stabilita relativních četností** možných výsledků, která s nezávislými opakováními pokusu roste
- **náhodný jev** tvrzení o výsledku náhodného pokusu (podmnožina množiny Ω)
- **jistý jev** Ω nastává vždy
- **nemožný jev** $\emptyset$ nenastává nikdy
- **podjev**: $B \subset D$ znamená $B \Rightarrow D$
- **jev opačný**: $\overline{D} \Leftrightarrow$ neplatí D
- **průnik jevů** $B \cap D$ nastaly oba jevy
- **sjednocení jevů** $D \cup B$ nastal **aspoň** jeden
- **neslučitelné jevy** $B \cap D = \emptyset$

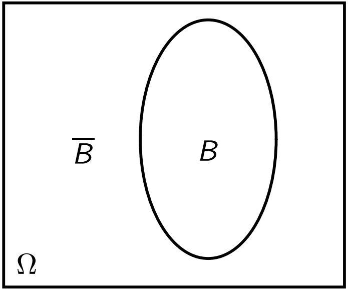
znázornění pomocí Vennova diagramu

celý obdélník – jev jistý

$B \subset D \Rightarrow \mathrm{P}(B) \leq \mathrm{P}(D)$



$\mathrm{P}(\overline{B}) = 1 - \mathrm{P}(B)$



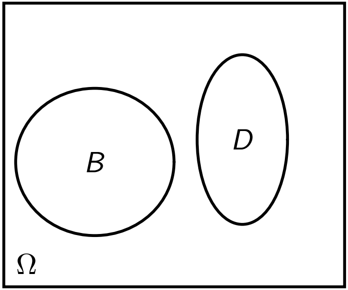
velikost plochy odpovídá pravděpodobnosti

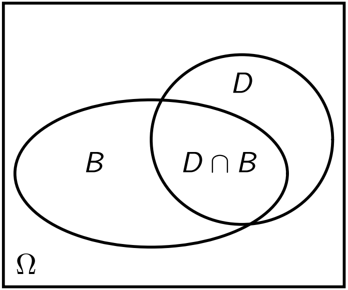
$$B \cap D = \emptyset \Rightarrow$$

$$P(B \cup D) = P(B) + P(D)$$

obecně platí

$$P(B \cup D) = P(B) + P(D) - P(B \cap D)$$





velikost plochy odpovídá pravděpodobnosti

pravděpodobnost

- objektivní číselné vyjádření „naděje“, že nastane jev B
- modelový protějšek relativní četnosti
- pravděpodobnost (pst) by měla mít stejné vlastnosti jako relativní četnost:
  - $\boxed{0 \leq \mathrm{P}(B) \leq 1}$
  - $\boxed{\mathrm{P}(\Omega) = 1, \mathrm{P}(\emptyset) = 0}$
  - $\boxed{B \cap D = \emptyset \Rightarrow \mathrm{P}(B \cup D) = \mathrm{P}(B) + \mathrm{P}(D)}$ **(sčítání pravděpodobností)**
  - $\boxed{\mathrm{P}(B \cup D) = \mathrm{P}(B) + \mathrm{P}(D) - \mathrm{P}(B \cap D)}$
  - $\boxed{B \subset D \Rightarrow \mathrm{P}(B) \leq \mathrm{P}(D)}$
  - $\boxed{\mathrm{P}(\overline{B}) = 1 - \mathrm{P}(B)}$

pravděpodobnost

- objektivní číselné vyjádření „naděje“, že nastane jev B
- modelový protějšek relativní četnosti
- pravděpodobnost (pst) by měla mít stejné vlastnosti jako relativní četnost:
  - $\boxed{0 \leq P(B) \leq 1}$
  - $\boxed{P(\Omega) = 1, P(\emptyset) = 0}$
  - $\boxed{B \cap D = \emptyset \Rightarrow P(B \cup D) = P(B) + P(D)}$
    (sčítání pravděpodobností)
  - $\boxed{P(B \cup D) = P(B) + P(D) - P(B \cap D)}$
  - $\boxed{B \subset D \Rightarrow P(B) \leq P(D)}$
  - $\boxed{P(\overline{B}) = 1 - P(B)}$

pravděpodobnost

- objektivní číselné vyjádření „naděje“, že nastane jev B
- modelový protějšek relativní četnosti
- pravděpodobnost (pst) by měla mít stejné vlastnosti jako relativní četnost:
  - $\boxed{0 \leq \mathrm{P}(B) \leq 1}$
  - $\boxed{\mathrm{P}(\Omega) = 1, \mathrm{P}(\emptyset) = 0}$
  - $\boxed{B \cap D = \emptyset \Rightarrow \mathrm{P}(B \cup D) = \mathrm{P}(B) + \mathrm{P}(D)}$ **(sčítání pravděpodobností)**
  - $\boxed{\mathrm{P}(B \cup D) = \mathrm{P}(B) + \mathrm{P}(D) - \mathrm{P}(B \cap D)}$
  - $\boxed{B \subset D \Rightarrow \mathrm{P}(B) \leq \mathrm{P}(D)}$
  - $\boxed{\mathrm{P}(\overline{B}) = 1 - \mathrm{P}(B)}$

pravděpodobnost

- objektivní číselné vyjádření „naděje“, že nastane jev B
- modelový protějšek relativní četnosti
- pravděpodobnost (pst) by měla mít stejné vlastnosti jako relativní četnost:
  - $\boxed{0 \leq \mathrm{P}(B) \leq 1}$
  - $\boxed{\mathrm{P}(\Omega) = 1, \mathrm{P}(\emptyset) = 0}$
  - $\boxed{B \cap D = \emptyset \Rightarrow \mathrm{P}(B \cup D) = \mathrm{P}(B) + \mathrm{P}(D)}$ **(sčítání pravděpodobností)**
  - $\boxed{\mathrm{P}(B \cup D) = \mathrm{P}(B) + \mathrm{P}(D) - \mathrm{P}(B \cap D)}$
  - $\boxed{B \subset D \Rightarrow \mathrm{P}(B) \leq \mathrm{P}(D)}$
  - $\boxed{\mathrm{P}(\overline{B}) = 1 - \mathrm{P}(B)}$

pravděpodobnost

- objektivní číselné vyjádření „naděje“, že nastane jev B
- modelový protějšek relativní četnosti
- pravděpodobnost (pst) by měla mít stejné vlastnosti jako relativní četnost:
  - $\boxed{0 \leq \mathrm{P}(B) \leq 1}$
  - $\boxed{\mathrm{P}(\Omega) = 1, \mathrm{P}(\emptyset) = 0}$
  - $\boxed{B \cap D = \emptyset \Rightarrow \mathrm{P}(B \cup D) = \mathrm{P}(B) + \mathrm{P}(D)}$
    (sčítání pravděpodobností)
  - $\boxed{\mathrm{P}(B \cup D) = \mathrm{P}(B) + \mathrm{P}(D) - \mathrm{P}(B \cap D)}$
  - $\boxed{B \subset D \Rightarrow \mathrm{P}(B) \leq \mathrm{P}(D)}$
  - $\boxed{\mathrm{P}(\overline{B}) = 1 - \mathrm{P}(B)}$

pravděpodobnost

- objektivní číselné vyjádření „naděje“, že nastane jev B
- modelový protějšek relativní četnosti
- pravděpodobnost (pst) by měla mít stejné vlastnosti jako relativní četnost:
  - $\boxed{0 \leq \mathrm{P}(B) \leq 1}$
  - $\boxed{\mathrm{P}(\Omega) = 1, \mathrm{P}(\emptyset) = 0}$
  - $\boxed{B \cap D = \emptyset \Rightarrow \mathrm{P}(B \cup D) = \mathrm{P}(B) + \mathrm{P}(D)}$ **(sčítání pravděpodobností)**
  - $\boxed{\mathrm{P}(B \cup D) = \mathrm{P}(B) + \mathrm{P}(D) - \mathrm{P}(B \cap D)}$
  - $\boxed{B \subset D \Rightarrow \mathrm{P}(B) \leq \mathrm{P}(D)}$
  - $\boxed{\mathrm{P}(\overline{B}) = 1 - \mathrm{P}(B)}$

pravděpodobnost

- objektivní číselné vyjádření „naděje“, že nastane jev B
- modelový protějšek relativní četnosti
- pravděpodobnost (pst) by měla mít stejné vlastnosti jako relativní četnost:
  - $\boxed{0 \leq \mathrm{P}(B) \leq 1}$
  - $\boxed{\mathrm{P}(\Omega) = 1, \mathrm{P}(\emptyset) = 0}$
  - $\boxed{B \cap D = \emptyset \Rightarrow \mathrm{P}(B \cup D) = \mathrm{P}(B) + \mathrm{P}(D)}$ **(sčítání pravděpodobností)**
  - $\boxed{\mathrm{P}(B \cup D) = \mathrm{P}(B) + \mathrm{P}(D) - \mathrm{P}(B \cap D)}$
  - $\boxed{B \subset D \Rightarrow \mathrm{P}(B) \leq \mathrm{P}(D)}$
  - $\boxed{\mathrm{P}(\overline{B}) = 1 - \mathrm{P}(B)}$

pravděpodobnost

- objektivní číselné vyjádření „naděje“, že nastane jev B
- modelový protějšek relativní četnosti
- pravděpodobnost (pst) by měla mít stejné vlastnosti jako relativní četnost:
  - $\boxed{0 \leq \mathrm{P}(B) \leq 1}$
  - $\boxed{\mathrm{P}(\Omega) = 1, \mathrm{P}(\emptyset) = 0}$
  - $\boxed{B \cap D = \emptyset \Rightarrow \mathrm{P}(B \cup D) = \mathrm{P}(B) + \mathrm{P}(D)}$ **(sčítání pravděpodobností)**
  - $\boxed{\mathrm{P}(B \cup D) = \mathrm{P}(B) + \mathrm{P}(D) - \mathrm{P}(B \cap D)}$
  - $\boxed{B \subset D \Rightarrow \mathrm{P}(B) \leq \mathrm{P}(D)}$
  - $\boxed{\mathrm{P}(\overline{B}) = 1 - \mathrm{P}(B)}$

pravděpodobnost

- objektivní číselné vyjádření „naděje“, že nastane jev B
- modelový protějšek relativní četnosti
- pravděpodobnost (pst) by měla mít stejné vlastnosti jako relativní četnost:
  - $\boxed{0 \leq \mathrm{P}(B) \leq 1}$
  - $\boxed{\mathrm{P}(\Omega) = 1, \mathrm{P}(\emptyset) = 0}$
  - $\boxed{B \cap D = \emptyset \Rightarrow \mathrm{P}(B \cup D) = \mathrm{P}(B) + \mathrm{P}(D)}$ **(sčítání pravděpodobností)**
  - $\boxed{\mathrm{P}(B \cup D) = \mathrm{P}(B) + \mathrm{P}(D) - \mathrm{P}(B \cap D)}$
  - $\boxed{B \subset D \Rightarrow \mathrm{P}(B) \leq \mathrm{P}(D)}$
  - $\boxed{\mathrm{P}(\overline{B}) = 1 - \mathrm{P}(B)}$

- **klasická definice psti**
  - m **stejně pravděpodobných** elementárních jevů $\omega_1, \ldots, \omega_m$
  - jsou neslučitelné, sjednocení všech je jistý jev
  - m_B elementárních jevů **příznivých jevu** B (tj. takových ω_i, že $\omega_i \in B$, je právě m_B)

$$\boxed{\mathrm{P}(B) = \frac{m_B}{m}}$$

- **příklad**
  - hází se dvěma kostkami (modrá, zelená)
  - B – součet aspoň 10

$$m = 6 \cdot 6 = 36; \quad m_B = 6 \quad \Rightarrow \quad \mathrm{P}(B) = \frac{6}{36}$$

příznivé možnosti: (6,4), (6,5), (6,6), (5,5), (5,6), (4,6)

- **klasická definice psti**
  - m **stejně pravděpodobných** elementárních jevů $\omega_1, \ldots, \omega_m$
  - jsou neslučitelné, sjednocení všech je jistý jev
  - m_B elementárních jevů **příznivých jevu** B (tj. takových ω_i, že $\omega_i \in B$, je právě m_B)

$$\boxed{\mathrm{P}(B) = \frac{m_B}{m}}$$

- **příklad**
  - hází se dvěma kostkami (modrá, zelená)
  - B – součet aspoň 10

$$m = 6 \cdot 6 = 36; \quad m_B = 6 \quad \Rightarrow \quad \mathrm{P}(B) = \frac{6}{36}$$

příznivé možnosti: (6,4), (6,5), (6,6), (5,5), (5,6), (4,6)

- **klasická definice psti**
  - m **stejně pravděpodobných** elementárních jevů $\omega_1, \ldots, \omega_m$
  - jsou neslučitelné, sjednocení všech je jistý jev
  - m_B elementárních jevů **příznivých jevu** B (tj. takových ω_i, že $\omega_i \in B$, je právě m_B)

$$\boxed{\mathrm{P}(B) = \frac{m_B}{m}}$$

- **příklad**
  - hází se dvěma kostkami (modrá, zelená)
  - B – součet aspoň 10

$$m = 6 \cdot 6 = 36; \quad m_B = 6 \quad \Rightarrow \quad \mathrm{P}(B) = \frac{6}{36}$$

příznivé možnosti: (6,4), (6,5), (6,6), (5,5), (5,6), (4,6)

- **klasická definice psti**
  - m **stejně pravděpodobných** elementárních jevů $\omega_1, \dots, \omega_m$
  - jsou neslučitelné, sjednocení všech je jistý jev
  - m_B elementárních jevů **příznivých jevu** B
    (tj. takových ω_i, že $\omega_i \in B$, je právě m_B)

$$\boxed{\mathrm{P}(B) = \frac{m_B}{m}}$$

- **příklad**
  - hází se dvěma kostkami (modrá, zelená)
  - B – součet aspoň 10

$$m = 6 \cdot 6 = 36; \quad m_B = 6 \quad \Rightarrow \quad \mathrm{P}(B) = \frac{6}{36}$$

příznivé možnosti: (6,4), (6,5), (6,6), (5,5), (5,6), (4,6)

- **klasická definice psti**
  - m **stejně pravděpodobných** elementárních jevů $\omega_1, \dots, \omega_m$
  - jsou neslučitelné, sjednocení všech je jistý jev
  - m_B elementárních jevů **příznivých jevu** B (tj. takových ω_i, že $\omega_i \in B$, je právě m_B)

$$\boxed{\mathrm{P}(B) = \frac{m_B}{m}}$$

- **příklad**
  - hází se dvěma kostkami (modrá, zelená)
  - B – součet aspoň 10

$$m = 6 \cdot 6 = 36; \quad m_B = 6 \quad \Rightarrow \quad \mathrm{P}(B) = \frac{6}{36}$$

příznivé možnosti: (6,4), (6,5), (6,6), (5,5), (5,6), (4,6)

- **klasická definice psti**
  - m **stejně pravděpodobných** elementárních jevů $\omega_1, \dots, \omega_m$
  - jsou neslučitelné, sjednocení všech je jistý jev
  - m_B elementárních jevů **příznivých jevu** B (tj. takových ω_i, že $\omega_i \in B$, je právě m_B)

$$\boxed{\mathrm{P}(B) = \frac{m_B}{m}}$$

- **příklad**
  - hází se dvěma kostkami (modrá, zelená)
  - B – součet aspoň 10

$$m = 6 \cdot 6 = 36; \quad m_B = 6 \quad \Rightarrow \quad \mathrm{P}(B) = \frac{6}{36}$$

příznivé možnosti: (6,4), (6,5), (6,6), (5,5), (5,6), (4,6)

- **klasická definice psti**
  - m **stejně pravděpodobných** elementárních jevů $\omega_1, \dots, \omega_m$
  - jsou neslučitelné, sjednocení všech je jistý jev
  - m_B elementárních jevů **příznivých jevu** B (tj. takových ω_i, že $\omega_i \in B$, je právě m_B)

$$\boxed{\mathrm{P}(B) = \frac{m_B}{m}}$$

- **příklad**
  - hází se dvěma kostkami (modrá, zelená)
  - B – součet aspoň 10

$$m = 6 \cdot 6 = 36; \quad m_B = 6 \quad \Rightarrow \quad \mathrm{P}(B) = \frac{6}{36}$$

příznivé možnosti: (6,4), (6,5), (6,6), (5,5), (5,6), (4,6)

hypergeometrické rozdělení

příklad na klasickou pravděpodobnost

- kombinační číslo $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1}$
- v rybníku je N ryb (zpravidla neznámý počet)
  A ryb vylovíme, označíme a vypustíme zpět
- po nějaké době vylovíme n ryb, z nich Y je označených
- s jakou pravděpodobností je $Y = k$?
  - celkem $\binom{N}{n}$ možných n-tic vylovených ryb
  - k označených lze vybrat $\binom{A}{k}$ způsoby
  - $n-k$ neoznačených lze vybrat $\binom{N-A}{n-k}$ způsoby

$$P(Y=k) = \frac{\binom{A}{k}\binom{N-A}{n-k}}{\binom{N}{n}}, \qquad \max(0, n+A-N) \le k \le \min(A, n)$$

- např. odhad neznámého N: $\hat{N} = n\frac{A}{Y}$ (neboť $Y/n \doteq A/N$)

hypergeometrické rozdělení

příklad na klasickou pravděpodobnost

- kombinační číslo $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1}$
- v rybníku je N ryb (zpravidla neznámý počet)
  A ryb vylovíme, označíme a vypustíme zpět
- po nějaké době vylovíme n ryb, z nich Y je označených
- s jakou pravděpodobností je $Y = k$?
  - celkem $\binom{N}{n}$ možných n-tic vylovených ryb
  - k označených lze vybrat $\binom{A}{k}$ způsoby
  - $n-k$ neoznačených lze vybrat $\binom{N-A}{n-k}$ způsoby

$$P(Y=k) = \frac{\binom{A}{k}\binom{N-A}{n-k}}{\binom{N}{n}}, \qquad \max(0, n+A-N) \le k \le \min(A, n)$$

- např. odhad neznámého N: $\hat{N} = n\frac{A}{Y}$ (neboť $Y/n \doteq A/N$)

hypergeometrické rozdělení
příklad na klasickou pravděpodobnost

- kombinační číslo $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1}$
- v rybníku je N ryb (zpravidla neznámý počet)
  A ryb vylovíme, označíme a vypustíme zpět
- po nějaké době vylovíme n ryb, z nich Y je označených
- s jakou pravděpodobností je $Y = k$?
  - celkem $\binom{N}{n}$ možných n-tic vylovených ryb
  - k označených lze vybrat $\binom{A}{k}$ způsoby
  - $n-k$ neoznačených lze vybrat $\binom{N-A}{n-k}$ způsoby

$$P(Y=k) = \frac{\binom{A}{k}\binom{N-A}{n-k}}{\binom{N}{n}}, \qquad \max(0, n+A-N) \le k \le \min(A, n)$$

- např. odhad neznámého N: $\hat{N} = n\frac{A}{Y}$ (neboť $Y/n \doteq A/N$)

hypergeometrické rozdělení

příklad na klasickou pravděpodobnost

- kombinační číslo $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1}$
- v rybníku je N ryb (zpravidla neznámý počet)
  A ryb vylovíme, označíme a vypustíme zpět
- po nějaké době vylovíme n ryb, z nich Y je označených
- s jakou pravděpodobností je $Y = k$?
  - celkem $\binom{N}{n}$ možných n-tic vylovených ryb
  - k označených lze vybrat $\binom{A}{k}$ způsoby
  - $n - k$ neoznačených lze vybrat $\binom{N-A}{n-k}$ způsoby

$$\mathrm{P}(Y = k) = \frac{\binom{A}{k}\binom{N-A}{n-k}}{\binom{N}{n}}, \qquad \max(0, n+A-N) \le k \le \min(A, n)$$

- např. odhad neznámého N: $\hat{N} = n\frac{A}{Y}$ (neboť $Y/n \doteq A/N$)

hypergeometrické rozdělení

příklad na klasickou pravděpodobnost

- kombinační číslo $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1}$
- v rybníku je N ryb (zpravidla neznámý počet)
  A ryb vylovíme, označíme a vypustíme zpět
- po nějaké době vylovíme n ryb, z nich Y je označených
- s jakou pravděpodobností je $Y = k$?
  - celkem $\binom{N}{n}$ možných n-tic vylovených ryb
  - k označených lze vybrat $\binom{A}{k}$ způsoby
  - $n - k$ neoznačených lze vybrat $\binom{N-A}{n-k}$ způsoby

$$P(Y = k) = \frac{\binom{A}{k}\binom{N-A}{n-k}}{\binom{N}{n}}, \qquad \max(0, n+A-N) \le k \le \min(A, n)$$

- např. odhad neznámého N: $\hat{N} = n\frac{A}{Y}$ (neboť $Y/n \doteq A/N$)

hypergeometrické rozdělení

příklad na klasickou pravděpodobnost

- kombinační číslo $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1}$
- v rybníku je N ryb (zpravidla neznámý počet)
  A ryb vylovíme, označíme a vypustíme zpět
- po nějaké době vylovíme n ryb, z nich Y je označených
- s jakou pravděpodobností je $Y = k$?
  - celkem $\binom{N}{n}$ možných n-tic vylovených ryb
  - k označených lze vybrat $\binom{A}{k}$ způsoby
  - $n-k$ neoznačených lze vybrat $\binom{N-A}{n-k}$ způsoby

$$P(Y = k) = \frac{\binom{A}{k}\binom{N-A}{n-k}}{\binom{N}{n}}, \qquad \max(0, n+A-N) \leq k \leq \min(A, n)$$

- např. odhad neznámého N: $\hat{N} = n\frac{A}{Y}$ (neboť $Y/n \doteq A/N$)

hypergeometrické rozdělení

příklad na klasickou pravděpodobnost

- kombinační číslo $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1}$
- v rybníku je N ryb (zpravidla neznámý počet)
  A ryb vylovíme, označíme a vypustíme zpět
- po nějaké době vylovíme n ryb, z nich Y je označených
- s jakou pravděpodobností je $Y = k$?
  - celkem $\binom{N}{n}$ možných n-tic vylovených ryb
  - k označených lze vybrat $\binom{A}{k}$ způsoby
  - $n - k$ neoznačených lze vybrat $\binom{N-A}{n-k}$ způsoby

$$\mathrm{P}(Y = k) = \frac{\binom{A}{k}\binom{N-A}{n-k}}{\binom{N}{n}}, \qquad \max(0, n+A-N) \le k \le \min(A, n)$$

- např. odhad neznámého N: $\hat{N} = n\frac{A}{Y}$ (neboť $Y/n \doteq A/N$)

hypergeometrické rozdělení

příklad na klasickou pravděpodobnost

- kombinační číslo $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1}$
- v rybníku je N ryb (zpravidla neznámý počet)
  A ryb vylovíme, označíme a vypustíme zpět
- po nějaké době vylovíme n ryb, z nich Y je označených
- s jakou pravděpodobností je $Y = k$?
  - celkem $\binom{N}{n}$ možných n-tic vylovených ryb
  - k označených lze vybrat $\binom{A}{k}$ způsoby
  - $n - k$ neoznačených lze vybrat $\binom{N-A}{n-k}$ způsoby

$$\mathrm{P}(Y = k) = \frac{\binom{A}{k}\binom{N-A}{n-k}}{\binom{N}{n}}, \qquad \max(0, n+A-N) \leq k \leq \min(A, n)$$

- např. odhad neznámého N: $\hat{N} = n\frac{A}{Y}$ (neboť $Y/n \doteq A/N$)

příklad rodina
(nejmladší, prostřední, nejstarší)

tři sourozenci, celkem 8 elementárních jevů $\omega_1, \ldots, \omega_8$

ω_i	D	B	$B \cap D$	$B \cup D$	C
(m, m, m)					+
(f, m, m)	+	+	+	+	+
(m, f, m)		+		+	+
(f, f, m)	+			+	+
(f, f, f)	+			+	
(m, f, f)					
(f, m, f)	+			+	
(m, m, f)		+		+	

D nejmladší je dívka, $\mathrm{P}(D) = 4/8 = 1/2$
B v rodině je jediná dívka, $\mathrm{P}(B) = 3/8$
$B \cap D$ jediná dívka je nejmladší, $\mathrm{P}(B \cap D) = 1/8$
$\mathrm{P}(B \cup D) = \mathrm{P}(B) + \mathrm{P}(D) - \mathrm{P}(B \cap D) = \frac{3}{8} + \frac{4}{8} - \frac{1}{8} = \frac{6}{8}$
C nejstarší je hoch, $\mathrm{P}(C) = 4/8 = 1/2$