# Cvičení z biostatistiky 04

Poslední úprava dokumentu: 13. března 2024.

## Klasická definice pravděpodobnosti a pravděpodobnostní rozdělení, zkoumání normality dat

Spusťte RStudio, otevřete si nový skript, uložte si ho a nastavte pracovní adresář.

S teorií k tomuto cvičení vám může pomoci text *Náhodná veličina a její rozdělení* (dále pod zkratkou *NVR*), který najdete na mých stránkách (`https://web.natur.cuni.cz/uamvt/turcm6am`), popřípadě si příslušné pojmy připomeňte z přednášky či z učebnice.

## 1 Klasická definice pravděpodobnosti

**Náhodný pokus** je pokus konaný za přesně definovaných podmínek, jehož výsledek je předem nejistý (např. hod kostkou). Nejjemněji rozlišované možné výsledky náhodného pokusu, které se vzájemně vylučují (nemohou nastat dva současně) a jeden vždy musí nastat, se nazývají **elementární jevy** (např. pro hod kostkou je množina elementárních jevů rovna $\{1, 2, 3, 4, 5, 6\}$). Množina všech elementárních jevů daného pokusu se obvykle značí písmenem $\Omega$. **Náhodný jev** $A$ je pak libovolné tvrzení o výsledku náhodného pokusu. Náhodné jevy značíme zpravidla velkými písmeny ze začátku abecedy. Korektně pak můžeme náhodný jev $A$ popsat tak, že vyjmenujeme, ze kterých elementárních jevů se skládá, tj. vyjmenujeme všechny elementární jevy, které jsou příznivé jevu $A$ (jev $A$, že na kostce padne sudé číslo, se skládá z elementárních jevů $\{2, 4, 6\}$).

Jsou-li všechny elementární jevy stejně pravděpodobné (což je příklad třeba spravedlivé kostky, kde všechny stěny mají pravděpodobnost 1/6), pak pravděpodobnost jevu $A$, který je složen z $m_A$ elementárních jevů, spočítáme jednoduše jako

$$P(A) = \frac{m_A}{m},$$

kde $m$ je celkový počet všech elementárních jevů (tj. celkový počet prvků množiny $\Omega$). Tímto vzorečkem se dříve pravděpodobnost definovala, a tak dnes říkáme, že jde o tzv. **klasickou definici pravděpodobnosti**. Například pro jev $A$, že na kostce padne sudé číslo je $m_A = 3$ a $m = 6$, a tudíž $P(A) = \frac{3}{6} = \frac{1}{2}$.

**1)** Kolika způsoby si může 15 studentů *Základů biostatistiky* sednout k 15 počítačům v učebně B5? Jaká je pravděpodobnost, že si sednou (v učebně zleva doprava) přesně v pořadí, v jakém jsou uvedeni v prezenční listině?

✧ Množina $\Omega$ je tvořena všemi permutacemi, které lze vytvořit z 15 prvků (studentů). Celkový počet těchto permutací je $m = 15!$. V R to spočítáme jako

```
factorial(15)
```

✧ Uspořádání, které vyhovuje prezenční listině, je pouze jediné. Tedy počet příznivých uspořádání je $m_A = 1$, a tudíž pravděpodobnost této jediné permutace je dle klasické definice pravděpodobnosti rovna

```
1/factorial(15)
```

**2)** Kolika způsoby si může sednout do posluchárny o 170 židlích 170 studentů?

✧ Vyjde vám nějaké ohromné číslo, protože faktoriál se zvětšuje velkou rychlostí.

**3)** Kolika způsoby si může sednout do posluchárny o 171 židlích 171 studentů?

✧ Zde už je odpověď `Inf` (= *infinity*, nekonečno). I R má své limity...

**4)** S jakou pravděpodobností vyhrajete ve Sportce 1. cenu s jedním vsazeným sloupcem?

✧ Ve Sportce se tipuje 6 ze 49 čísel. Vyhrát první cenu znamená uhodnout všech šest čísel.

✧ Princip losování přitom zaručuje, že každá šestice čísel má stejnou pravděpodobnost. Množina $\Omega$ je tedy tvořena všemi šesticemi, které lze vytvořit ze 49 čísel. Těchto šestic je $\binom{49}{6}$, což v R vypočteme jako

```
choose(49, 6)
```

✧ Počet příznivých jevů, tj. počet šestic, na něž připadne první cena, je roven 1. Proto pravděpodobnost výhry první ceny je dle klasické definice pravděpodobnosti rovna

```
1 / choose(49, 6)
```

**5)** S jakou pravděpodobností vyhrajete ve Sportce 1. pořadí s úplně vyplněným tiketem (10 sloupců), jestliže v každém sloupci máte uvedenu jinou kombinaci vsazených čísel?

```
10 / choose(49, 6)
```

**6)** Házíme dvěma kostkami. S jakou pravděpodobností bude:

- ✧ součet obou hodnot roven 10?
- ✧ součet obou hodnot roven alespoň 10?

## 2 Výpočet střední hodnoty a rozptylu z definice

Připomeňte si vzorec pro výpočet střední hodnoty diskrétní náhodné veličiny (např. *NVR*, str. 3, vzorec (1)).

### Příklad

Náhodná veličina $X$ nabývá pouze hodnot $2, 3, 5$ a to s pravděpodobnostmi

$$P(X = 2) = 0.6 \qquad P(X = 3) = 0.3 \qquad P(X = 5) = 0.1.$$

Vypočtětě střední hodnotu a rozptyl této náhodné veličiny.

✧ **střední hodnotu** vypočteme z definice jako

$$\begin{aligned} \mathrm{E}\,X &= 2 \cdot P(X = 2) + 3 \cdot P(X = 3) + 5 \cdot P(X = 5) \\ &= 2 \cdot 0.6 + 3 \cdot 0.3 + 5 \cdot 0.1 = 2.6 \end{aligned}$$

✧ **rozptyl** můžeme také vypočítat z definice (viz např. *NVR*, str. 4, vzorec (2))

$$\begin{aligned} \operatorname{var} X &= \mathrm{E}(X - \mathrm{E}\,X)^2 \\ &= (2 - \mathrm{E}\,X)^2 \cdot P(X = 2) + (3 - \mathrm{E}\,X)^2 \cdot P(X = 3) + (5 - \mathrm{E}\,X)^2 \cdot P(X = 5) \\ &= (2 - 2.6)^2 \cdot 0.6 + (3 - 2.6)^2 \cdot 0.3 + (5 - 2.6)^2 \cdot 0.1 \\ &= 0.84 \end{aligned}$$

✧ Někdy je výhodnější vypočítat rozptyl podle vzorečku

$$\operatorname{var} X = \mathrm{E}(X^2) - (\mathrm{E}\, X)^2$$

K tomu je potřeba si vypočítat $\mathrm{E}(X^2)$ (tzv. druhý moment veličiny $X$), což uděláme opět z definice střední hodnoty (tentokrát jde o střední hodnotu $X^2$)

$$\begin{aligned}\mathrm{E}(X^2) &= 2^2 \cdot P(X=2) + 3^2 \cdot P(X=3) + 5^2 \cdot P(X=5) \\ &= 4 \cdot 0.6 + 9 \cdot 0.3 + 25 \cdot 0.1 = 7.6\end{aligned}$$

a pak dosadit

$$\operatorname{var} X = \mathrm{E}(X^2) - (\mathrm{E}\, X)^2 = 7.6 - (2.6)^2 = 0.84.$$

**Příklad - samostatná práce**

Náhodná veličina $X$ nabývá pouze hodnot 0 a 1, a to s pravděpodobnostmi

$$P(X=0) = 0.8 \qquad\qquad P(X=1) = c.$$

Určete hodnotu $c$ a spočtěte $\mathrm{E}\, X$.

## 3 Binomické rozdělení

Připomeňte si význam veličiny s binomickým rozdělením (např. *NVR* str. 7) a základní charakteristiky tohoto rozdělení. Máme posloupnost $n$ nezávislých dílčích pokusů, v nichž náhodný jev *zdar* nastává s pravděpodobností $p$. Náhodná veličina s binomickým rozdělením s parametry $n, p$ je dána celkovým počtem zdarů v takové posloupnosti. Platí:

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad \mathrm{E}\, X = np, \qquad \operatorname{var} X = np(1-p). \tag{1}$$

**Příklad s krajtou**

Předpokládejme, že se z vajíčka krajty vylíhne živý jedinec s pravděpodobností 0.5. V následujících úkolech nás bude zajímat pravděpodobnost, s jakou se z 10 vajíček vylíhne určitý počet živých krajt. Veličinou, která nás zajímá je tedy počet úspěchů (= živých krajt) z 10 pokusů (= vajíček). To je přesně veličina s binomickým rozdělením a parametry $n = 10$ a $p = 0.5$. Označíme-li si tuto veličinu jako $X$, pak můžeme symbolicky psát $X \sim Bi(10, 0.5)$.

**1)** S jakou pravděpodobností se z 10 vajíček vylíhne **právě** 5 živých krajt?

✧ Zajímá nás $P(X=5)$. Dle vzorečku (1) víme, že je to

$$P(X=5) = \binom{10}{5} (0.5)^5 (1-0.5)^{(10-5)} = \binom{10}{5} (0.5)^5 (0.5)^5,$$

✧ což v R můžeme jednoduše spočítat pomocí

```
dbinom(5, 10, 0.5)
```

✧ Odkud se vzal název příkazu? Zkratka `binom` v příkazu samozřejmě odkazuje na binomické rozdělení, předpona `d` pak zkracuje *density* (hustota), neboť pravděpodobnosti $P(X=k)$ jsou vlastně analogií hustoty u spojitých rozdělení.

**2)** S jakou pravděpodobností se z 10 vajíček vylíhne **nejvýše** 5 živých krajt?

✧ Tady je dotaz na $P(X \leq 5)$, což je hodnota distribuční funkce binomického rozdělení v bodě 5. Z teorie víme, že

$$P(X \leq 5) = \sum_{k=0}^{5} P(X = k) = \sum_{k=0}^{5} \binom{10}{k} (0.5)^k (1 - 0.5)^{(10-k)},$$

✧ což v R vypočteme jako

```
pbinom(5, 10, 0.5)
```

(Předpona `p` v názvu příkazu odkazuje na *probability*, ale nedá se říct, že by toto označení bylo nějak zvlášť šťastné...)

**3)** S jakou pravděpodobností se z 10 vajíček vylíhne **alespoň** 5 živých krajt?

✧ Zde je dotaz na $P(X \geq 5)$, což můžeme jednoduše převést na předchozí případ pomocí

$$P(X \geq 5) = 1 - P(X \leq 4).$$

✧ V R to pak vypočteme jako

```
1 - pbinom(4, 10, 0.5)
```

**4)** Všechny pravděpodobnosti $P(X = k)$ pro $k = 0, 1, \ldots, 10$ lze vypsat příkazem:

```
dbinom(0:10, 10, 0.5)
```

**5)** Nechte si graficky znázornit pravděpodobnosti jednotlivých hodnot. Dále si nechte vykreslit distribuční funkci počtu vylíhnutých krajt a připomeňte si, jaké vlastnosti má distribuční funkce diskrétního rozdělení.

✧ Jednotlivé pravděpodobnosti si lze vykreslit pomocí příkazů:

```
kk <- 0:10 # možné hodnoty
cbind("k"=kk,"pst"=round(dbinom(kk,10,1/2),6))
plot(kk,dbinom(kk,10,1/2),pch=16)
for (k in kk) lines(c(k,k),c(0,dbinom(k,10,1/2)))
abline(h=0,col="grey")
```

✧ Distribuční funkci vykreslíme příkazem:

```
plot(stepfun(kk,c(0,pbinom(kk,10,1/2))),pch=16,verticals=FALSE,
main="Distribuční funkce")
abline(h=0:1,col="grey")
```

### Příklad - kostka

Házíme 20-krát symetrickou kostkou. Zajímá nás jaká je pravděpodobnost, že z těchto 20 hodů padně určitý počet šestek. Rozmyslete si, že veličina $X$ představující počet šestek ve 20 hodech má binomické rozdělení $Bi(20, \frac{1}{6})$.

✧ S jakou pravděpodobností padnou právě 4 šestky?

✧ S jakou pravděpodobností padnou nejvýše 4 šestky?

✧ S jakou pravděpodobností padne alespoň 5 šestek?

# 4 Poissonovo rozdělení

Na rozdíl od binomického rozdělení zde není konečný počet dílčích pokusů, v nichž bychom zjišťovali, zda nastal či nenastal sledovaný náhodný jev (*zdar*). Zajímá nás počet výskytů sledovaného náhodného jevu během daného časového intervalu či na dané ploše „jednotkové“ velikosti apod., přičemž počty výskytů v různých časových intervalech (plochách) jsou nezávislé. Možné hodnoty nejsou shora nijak ohraničené, i když jsou velké hodnoty málo pravděpodobné. Střední hodnota i rozptyl jsou stejné, rovné jedinému parametru $\lambda > 0$. Platí tedy:

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad \mathrm{E}\, X = \lambda, \qquad \mathrm{var}\, X = \lambda. \tag{2}$$

**6) Příklad - studenti**: Předpokládejme, že se počet studentů, kteří dorazí na přednášku v intervalu 0-10 minut před jejím začátkem, řídí Poissonovým rozdělením s parametrem $\lambda = 15$.

✧ Připomeňte si, jak vypadá Poissonovo rozdělení. Nechte si znázornit pravděpodobnosti jednotlivých hodnot pro tento příklad.

```
lambda <- 15
kk <- 0:50       # omezíme se na tyto možné hodnoty
dpois(kk,lambda)
plot(kk,dpois(kk,lambda),pch=16)
arrows(x0=kk,y0=0,x1=kk,y1=dpois(kk,lambda),length=0)
abline(h=0,col="grey")
```

✧ Který počet studentů je nejpravděpodobnější?

  ✧ Nejpravděpodobnější je počet 14 a 15 studentů.

✧ Spočtěte pravděpodobnost, že na přednášce bude právě 10 studentů.

- Výpočet proveďte „ručně“ pomocí vzorečku.

```
lambda^10/factorial(10)*exp(-lambda)
```

- K výpočtu použijte funkci dpois(k,lambda) pro dané k a $\lambda$.

```
dpois(10,lambda)
```

✧ Určete pravděpodobnost, s jakou na přednášku nikdo nedorazí.

```
dpois(0,lambda)
```

✧ Jaká je pravděpodobnost, že dorazí maximálně pět studentů?

```
ppois(5,lambda)
```

✧ S jakou pravděpodobností nebude stačit posluchárna s kapacitou 25 míst (ve smyslu, že někdo nebude mít místo k sezení).

```
1-ppois(25,lambda)
```

### Poznámka - Poissonovo rozdělení

V textu *NVR*, str. 8, si přečtěte sekci o Poissonovu rozdělení a o možnosti jeho využití k aproximaci binomického rozdělení.

Porovnejte graficky rozdělení $Bi(n,p)$ a $Po(np)$ pro různá $n$ a malá $p$. Uvažujte například $n = 20, p = 1/10$, dále $n = 50, p = 1/25$ a nakonec $n = 100, p = 1/50$. Uvidíte, že rozdělení $Bi(n,p)$ lze pro malá $p$ velmi dobře aproximovat pomocí $Po(np)$.

```
par(mfrow=c(2,1))   # umozni mit dva obrazky pod sebou
n <- 50; p <- 1/25
kk <- 0:n
```

```
plot(kk,dbinom(kk,n,p),pch=16,ylab="Binom")
arrows(x0=kk,y0=0,x1=kk,y1=dbinom(kk,n,p),length=0)
abline(h=0,col="grey")
plot(kk,dpois(kk,n*p),pch=16, ylab="Poiss")
arrows(x0=kk,y0=0,x1=kk,y1=dpois(kk,n*p),length=0)
abline(h=0,col="grey")
par(mfrow=c(1,1))   # aby byl v grafickem okne opet pouze 1 obrazek
```

# 5 Normální rozdělení

V textu *NVR*, str. 10, si přečtěte sekci o normálním rozdělení a jeho charakteristikách.

Na normálním rozdělení je mimo jiné zajímavé i to, že má zvlášť parametr pro střední hodnotu a zvlášť parametr pro rozptyl. Jeho střední hodnota a rozptyl spolu tedy nejsou žádným způsobem provázané. To u jiných rozdělení nenajdeme.

### Vykreslete si „gaussovku“

Hustota normálního rozdělení (slangově nazývaná „gaussovka“) je funkce daná předpisem

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{3}$$

kde $\mu$ a $\sigma^2$ jsou střední hodnota a rozptyl, kteréžto jsou parametry normálního rozdělení.

Vykreslete si „gaussovku“ pro různé hodnoty střední hodnoty a rozptylu (resp. směrodatné odchylky) a pozorujte vliv těchto parametrů na její tvar.

```
strhod = 1        # stredni hodnota, zkousejte ruzne hodnoty
smerodch = 3       # smerodatna odchylka, zkousejte ruzne hodnoty
curve(dnorm(x,strhod,smerodch),xlim=c(-30,30))
```

✧ **Pro zájemce:** Příkaz `dnorm` poskytuje hodnotu hustoty normálního rozdělení v daném bodě s danými parametry. Předpona `d` je zkrácením slova *density* (hustota). Prvním argumentem funkce `dnorm` je bod, ve kterém nás hustota zajímá (zde $x$, neboť nás to zajímá ve všech bodech a funkce `curve` přijímá pouze funkce proměnné $x$). Druhým argumentem je střední hodnota a třetím argumentem je směrodatná odchylka.

✧ Příkaz `curve` dokáže vykreslit libovolnou křivku proměnné $x$. Jeho argument `xlim` pak udává limity na ose x, v nichž má být křivka vykreslena.

✧ Jak jste zpozorovali, střední hodnota normálního rozdělení udává střed hustoty (tj. střed „kopečku“). Rozptyl pak udává, jak bude křivka „roztažená“ do stran. Možná jste si také všimli, že při větší hodnotě rozptylu se výška kopečku sníží (hodnoty na ose $y$ klesnou). Je to proto, že celková plocha pod hustotou musí být vždy rovna jedné, a tudíž je nutno tuto plochu i při větším rozptylu (a tedy „roztaženější“ křivce) zachovat.

### Příklad - IQ

Hodnota IQ koeficientu je tabelována tak, aby bylo jeho rozdělení normální se střední hodnotou 100 a směrodatnou odchylkou 15 (tímto tiše předpokládáme, že každý člověk je charakterizován hodnotou IQ určenou s neomezenou přesností).

Označme si veličinu představující IQ náhodně vybraného člověka jako $X$. Pak $X \sim N(100, 15^2)$. (Druhým parametrem normálního rozdělení je tradičně rozptyl, nikoli směrodatná odchylka, proto musíme 15 umocnit na druhou).

**1)** Určeme pravděpodobnost, že IQ náhodně vybraného člověka je nižší než 120.

✧ Zajímá nás tedy $P(X \leq 120)$, což je hodnota distribuční funkce normálního rozdělení $N(100, 15^2)$ v bodě 120. Žádný explicitní vzorec pro distribuční funkci normálního rozdělení bohužel neexistuje, lze ji zapsat pouze pomocí integrálu z hustoty, tj.

$$P(X \leq z) = \int_{-\infty}^{z} f(x)dx = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx,$$

kde $f(x)$ je hustota normálního rozdělení $N(\mu, \sigma^2)$, kterou jsme již viděli ve vzorci (3). Pro nás tedy

$$P(X \leq 120) = \int_{-\infty}^{120} \frac{1}{\sqrt{2\pi 15^2}} e^{-\frac{(x-100)^2}{2 \cdot 15^2}} dx.$$

✧ Naštěstí tento integrál nemusíme počítat ručně, ale spočítá ho za nás R:

```
pnorm(120, 100, 15)
```

**2)** Určeme pravděpodobnost, že IQ náhodně vybraného člověka je vyšší než 120.

✧ Zde je dotaz na $P(X > 120)$, což můžeme opět spočítat jako:

$$P(X > 120) = 1 - P(X \leq 120),$$

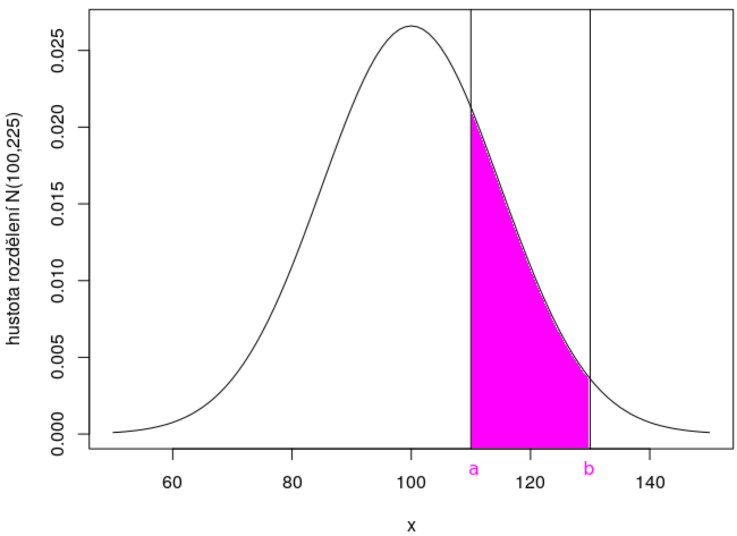
✧ což v R spočteme jako

```
1 - pnorm(120, 100, 15)
```

**3)** Určete pravděpodobnost, že IQ náhodně vybraného člověka leží v intervalu (110, 130).

✧ Pravděpodobnost, že veličina $X$ nabyde hodnoty z nějakého intervalu $(a, b)$ je rovna ploše po křivkou hustoty v tomto intervalu. A tato plocha se vypočte jakožto určitý integrál z hustoty $f(x)$ přes interval $(a, b)$:

$$P(X \in (a, b)) = P(a < X < b) = \int_{a}^{b} f(x)dx.$$



- Místo počítání integrálu si ale můžeme rozmyslet, že

  $$P(a < X < b) = P(X < b) - P(X < a),$$

  tedy jde o rozdíl distribuční funkce v bodě $b$ a v bodě $a$.
- V našem případě je $a = 110$ a $b = 130$. Rozdíl distribučních funkcí pak v R spočítáme jako

  ```
  pnorm(130, 100, 15) - pnorm(110, 100, 15)
  ```
- Možná vás zarazilo, že distribuční funkcí zde nazýváme pravděpodobnost $P(X < z)$, místo tradičního $P(X \leq z)$. Pravda je, že u spojitých rozdělení je to jedno. Je to díky tomu, že u spojitých rozdělení je pravděpodobnost jednoho konkrétního bodu rovna nule, tj. $P(X = z) = 0$. A tudíž $P(X \leq z) = P(X = z) + P(X < z) = P(X < z)$. Podrobněji se o tom můžete dočíst níže.

**4)** Určete pravděpodobnost, že IQ náhodně vybraného člověka je rovno 120.

- Správná odpověď je

  $$P(X = 120) = 0.$$

  Někdo bude možná zkoušet: `dnorm(120, 100, 15)`
  ALE POZOR! Pro spojitá rozdělení je P$(X = x) = 0$ ! Hustota je pro spojitá rozdělení sice jakousi „náhražkou“ za výraz P$(X = x)$ (který se vyskytuje u rozdělení diskrétních), avšak má význam pouze při určení pravděpodobnosti, že náhodná veličina $X$ leží v nějakém intervalu (tato pravděpodobnost je pak rovna ploše pod křivkou hustoty v daném intervalu). Hustotu tedy nelze použít pro výpočet pravděpodobnosti, že se $X$ rovná určité hodnotě, tj. $f(x) \neq P(X = x)$. Příkaz dnorm(120,100,15) nám sice dá nějaké číslo (neboť hustota $f(x)$ je funkcí proměnné $x$, takže $f(120)$ je „nějaké číslo“), avšak toto číslo nemá význam pravděpodobnosti jevu „IQ=120“, tj. $f(120) \neq P(\text{IQ} = 120)$.

## 6 Zkoumání normality dat

Mnohé metody, s nimiž se ještě potkáme, předpokládají u kvantitativních dat, že odpovídají normálnímu (Gaussovu) rozdělení. Toto rozdělení odvodil Abraham de Moivre v roce 1733, ale později dostalo název po německém matematikovi C. F. Gaussovi.


Obrázek 1: Německý matematik a fyzik Carl Friedrich Gauss (1777 - 1855) na staré německé marce.

**1)** Načtěte si opět datovou tabulku z minulé hodiny:

```
load("Kojeni.RData")
attach(Kojeni)
```

### QQ plot

Shodu rozdělení dat s předpokládaným rozdělením lze graficky nejlépe posoudit pomocí kvantil-kvantilového grafu (angl. *QQ plot*, což je zkratka pro *quantile-quantile plot*). Chceme-li tedy posoudit, zda by naše data mohla pocházet z normálního rozdělení, je nejlepším obrázek právě QQ plot, v případě normálního rozdělení nazývaný *normální diagram*.

V normálním diagramu jsou na ose $x$ teoretické kvantily normálního rozdělení $N(0,1)$ (ty jsou známy) a na ose $y$ jsou výběrové kvantily vypočítané z příslušných dat (výběrové kvantily jsme počítali na předchozích cvičeních pomocí příkazu `quantile`). Počet vykreslených kvantilů je roven počtu našich dat. Pocházejí-li naše data z normálního rozdělení (ne nutně jen $N(0,1)$), měly by si teoretické a výběrové kvantily odpovídat a body v normálním diagramu by měly ležet na přímce. Pokud data z normálního rozdělení nepocházejí, budou se body v normálním diagramu kolem přímky různě vlnit, popř. utíkat na koncích.

Nenechte se zmást tím, že výběrové kvantily v normálním diagramu porovnáváme s teoretickými kvantily zrovna $N(0,1)$ rozdělení. Pokud by naše data pocházela z normálního rozdělení s nějakou jinou střední hodnotou a jiným rozptylem (tj. obecně $N(\mu,\sigma^2)$), bude mít přímka z bodů akorát jiný sklon, ale její tvar zůstane zachován.

Posouzení normality na základě normálního diagramu je samozřejmě zcela subjektivní. Později budeme normální rozdělení dat rigorózně testovat pomocí **Shapiro-Wilkova testu**.

**2)** Nakreslete **normální diagram** pro věk matek (vek.m):

```
qqnorm(vek.m)
qqline(vek.m)
```

✧ Ukazuje obrázek na nenormalitu rozdělení věku matek?

**3)** Obrázek samozřejmě můžete dále zkrášlovat:

```
qqnorm(vek.m, pch=16, col="red", xlab="Teoreticke kvantily N(0, 1)",
        ylab="Kvantily veku matek", main="Normalni QQ graf (vek matek)")
qqline(vek.m, lty=6, col="darkblue")
```

## 7 Konec práce

Než zavřete všechna okna, nezapomeňte si uložit poslední změny ve skriptovém souboru:
File ➥ Save