

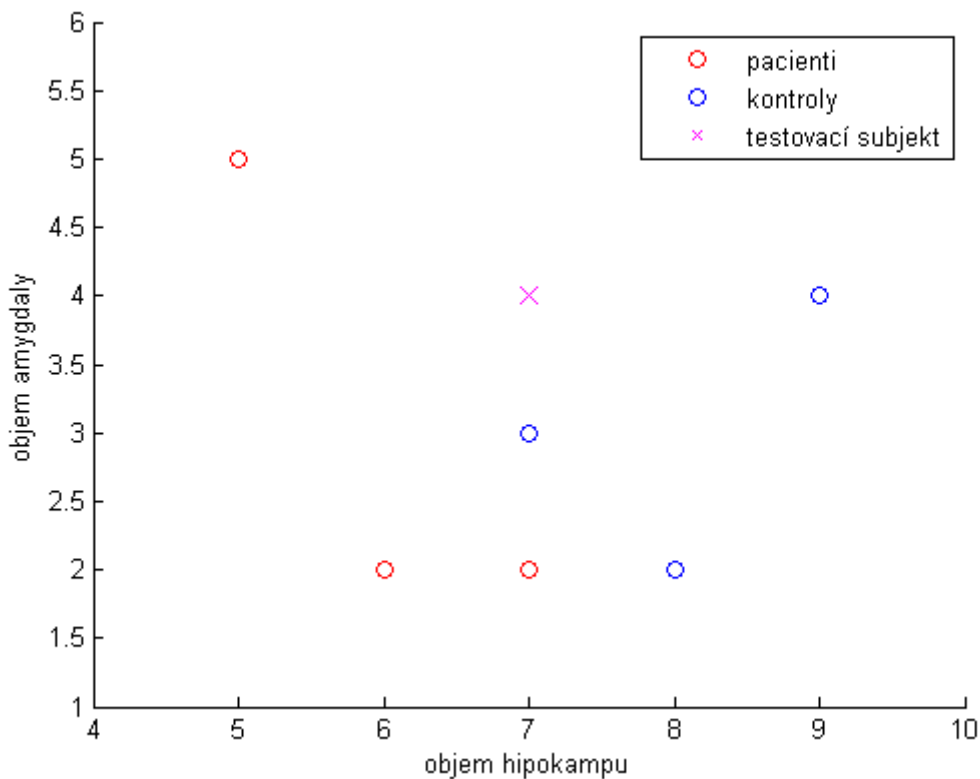
Bayesův klasifikátor – příklad

Bylo provedeno měření objemu hipokampu a amygdaly (v cm^3) u 3 pacientů s Alzheimerovou chorobou () a 3 kontrolních subjektů () (označení D – diseased, H – healthy). Naměřené hodnoty objemu hipokampu a amygdaly u pacientů (resp.) a kontrol (resp.) byly zaznamenány do matic resp. :

Určete, zda testovací subjekt patří do skupiny pacientů či kontrolních subjektů pomocí Bayesova klasifikátoru.

Řešení:

Nejprve si data znázorníme (Obrázek 1).



Obrázek 1. Vizualizace dat pacientů, kontrolních subjektů a testovacího subjektu.

Vyjdeme z Bayesova vzorce: $P(A|B) = \frac{P(A)P(B|A)}{P(B)}$, kde $P(A|B)$ je aposteriorní pravděpodobnost, $P(B|A)$ je podmíněná hustota pravděpodobnosti výskytu obrazu B ve třídě A ; $P(A)$ je apriorní pravděpodobnost třídy A a $P(B)$ je celková hustota pravděpodobnosti rozložení obrazu B v celém obrazovém prostoru.

Nejprve vypočteme apriorní pravděpodobnosti třídy pacientů a kontrol: $P(A) = \frac{1}{n_A}$ a $P(B) = \frac{1}{n_B}$

Dále vypočteme vícerozměrné průměry \bar{x}_A , \bar{x}_B a kovarianční matice S_A , S_B (předpokládáme, že data mají vícerozměrné normální rozdělení).

Vícerozměrné průměry pro třídu pacientů a kontrol:

$$\bar{x}_A = \begin{pmatrix} \bar{x}_{A1} \\ \bar{x}_{A2} \\ \bar{x}_{A3} \\ \bar{x}_{A4} \end{pmatrix}, \quad \bar{x}_B = \begin{pmatrix} \bar{x}_{B1} \\ \bar{x}_{B2} \\ \bar{x}_{B3} \\ \bar{x}_{B4} \end{pmatrix}$$

Výběrové kovarianční matice: S_A a S_B . Dílčí výpočty jednotlivých prvků výběrových kovariančních matic:

Rozptyl objemu hipokampu u pacientů:

$$s_{A1}^2 = \frac{1}{n_A} \sum_{i=1}^{n_A} (x_{i1} - \bar{x}_{A1})^2 = \frac{1}{12} (6-6)^2 + \frac{1}{12} (6-6)^2 + \frac{1}{12} (7-6)^2 + \frac{1}{12} (6-6)^2 = 1$$

Druhý způsob výpočtu: $s_{A1}^2 = \frac{1}{n_A} \sum_{i=1}^{n_A} x_{i1}^2 - \bar{x}_{A1}^2 = \frac{1}{12} (6^2 + 6^2 + 7^2 + 6^2) - 6^2 = 121 + 0 + 1 = 1$

Rozptyl objemu amygdaly u pacientů: $s_{A2}^2 = \frac{1}{n_A} \sum_{i=1}^{n_A} (x_{i2} - \bar{x}_{A2})^2 = \frac{1}{12} (13-15)^2 + \frac{1}{12} (15-15)^2 + \frac{1}{12} (65-15)^2 + \frac{1}{12} (3-15)^2 = 12 - 2 + 0 - 1 = -1,5$

Kovariance objemu hipokampu a objemu amygdaly u pacientů: $s_{A12} = \frac{1}{n_A} \sum_{i=1}^{n_A} (x_{i1} - \bar{x}_{A1})(x_{i2} - \bar{x}_{A2}) = \frac{1}{12} (6-6)(13-15) + \frac{1}{12} (6-6)(15-15) + \frac{1}{12} (7-6)(65-15) + \frac{1}{12} (6-6)(3-15) = 13 - 15 - 65 - 3 + 6 - 62 - 3 + 7 - 62 - 3 = 12 - 2 + 0 - 1 = -1,5$

Rozptyl objemu hipokampu u kontrol: $s_{B1}^2 = \frac{1}{n_B} \sum_{i=1}^{n_B} (x_{i1} - \bar{x}_{B1})^2 = \frac{1}{12} (13-15)^2 + \frac{1}{12} (15-15)^2 + \frac{1}{12} (65-15)^2 + \frac{1}{12} (3-15)^2 = 12 - 2 + 0 - 1 = -1,5$

Rozptyl objemu amygdaly u kontrol: _____

_____ -

Kovariance objemu hipokampu a objemu amygdaly u kontrol: _____

$$x_2 + x_{21} - x_1 \quad x_{22} - x_2 + x_{31} - x_1 \quad x_{32} - x_2 = 13 - 19 - 84 - 3 + 7 - 83 - 3 + 8 - 82 - 3 = 121 + 0 + 0 = 0,5$$

Výběrové kovarianční matice: a .

Pokud bychom kovarianční matici pacientů chtěli spočítat maticově: _____

$$x_{11} - x_1 \quad x_{12} - x_2 + x_{21} - x_1 \quad x_{22} - x_2 \quad x_{21} - x_1 \quad x_{22} - x_2 + x_{31} - x_1 \quad x_{32} - x_2 \quad x_{31} - x_1 \quad x_{32} - x_2 = 13 - 15 - 65 - 3 \cdot 5 - 65 - 3 + 6 - 62 - 3 \cdot 6 - 62 - 3 + 7 - 62 - 3 \cdot 7 - 62 - 3 = 12 - 12 \cdot -12 + 0 - 1 \cdot 0 - 1 + 1 - 1 \cdot 1 - 1 = 121 - 2 - 24 + 0001 + 1 - 1 - 11 = 1 - 1,5 - 1,53$$

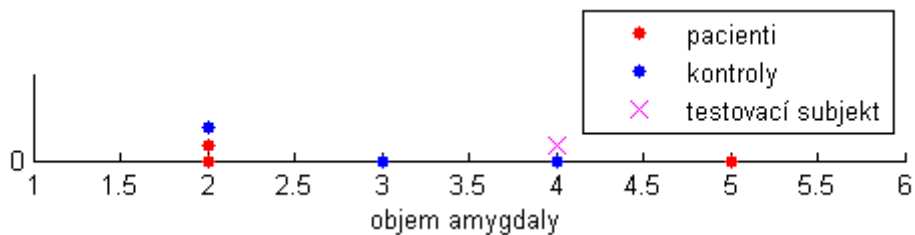
Na závěr vypočteme výběrový (Pearsonův) korelační koeficient objemu hipokampu a amygdaly u pacientů () a kontrolních subjektů ():

=====
 =====

Kritérium maximální aposteriorní pravděpodobnosti

1. Klasifikace podle objemu amygdaly:

Nejprve si znázorníme objem amygdaly u jednotlivých subjektů (Obrázek 2).



Obrázek 2. Vizualizace objemu amygdaly u jednotlivých subjektů.

Spočteme aposteriorní pravděpodobnosti _____ a _____ s využitím Bayesova vzorce, tudíž _____ a _____, a zařadíme testovací subjekt do třídy s větší aposteriorní pravděpodobností.

Výpočet _____, podmíněné hustoty pravděpodobnosti výskytu obrazu _____ ve třídě _____, a _____, podmíněné hustoty pravděpodobnosti výskytu obrazu _____ ve třídě _____ (grafické znázornění podmíněných hustot pravděpodobnosti viz *Obrázek 3*):

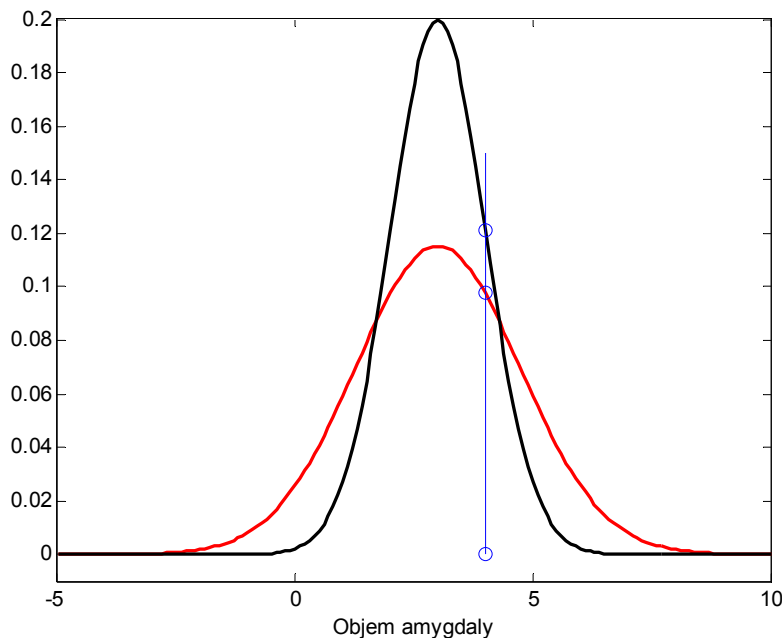
$$\frac{\text{_____}}{\text{_____}} = \frac{\text{_____}}{\text{_____}}$$

Výpočet celkové hustoty pravděpodobnosti:

Aposterioorní pravděpodobnosti: _____ a _____

(tzn. s pravděpodobností 44,6% bude subjekt zařazen do třídy pacientů a s pravděpodobností 55,4% do třídy kontrolních subjektů). Protože _____, zařadíme testovací subjekt do třídy kontrolních subjektů.

Poznámka: součet aposteriorních pravděpodobností je roven 1.

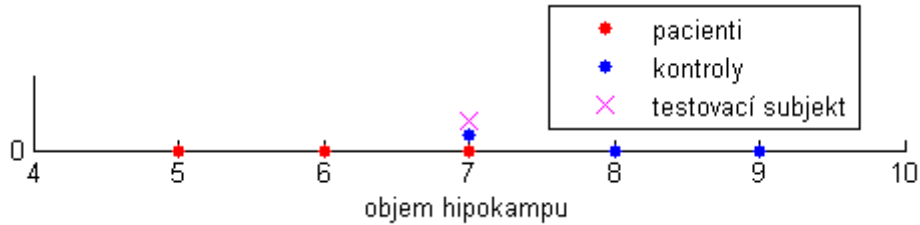


Obrázek 3. Vizualizace hustoty pravděpodobnosti pacientů (znázorněna červeně) a kontrolních subjektů (znázorněna černě). Podmíněné hustoty pravděpodobnosti výskytu testovacího subjektu _____ v jednotlivých třídách jsou znázorněny modře. Je patrné, že subjekt bude zařazen do třídy pacientů.

2. Klasifikace podle objemu hipokampu:

Spočteme aposteriorní pravděpodobnosti _____ a _____ s využitím Bayesova vzorce, tudíž _____ a _____, a zařadíme testovací subjekt do třídy s větší aposteriorní pravděpodobností.

Nejprve si znázorníme objem hipokampu u jednotlivých subjektů (Obrázek 4).



Obrázek 4. Vizualizace objemu hipokampu u jednotlivých subjektů.

Výpočet podmíněné hustoty pravděpodobnosti výskytu obrazu ve třídě _____ :

$$\frac{\text{_____}}{\text{_____}} = \frac{\text{_____}}{\text{_____}}$$

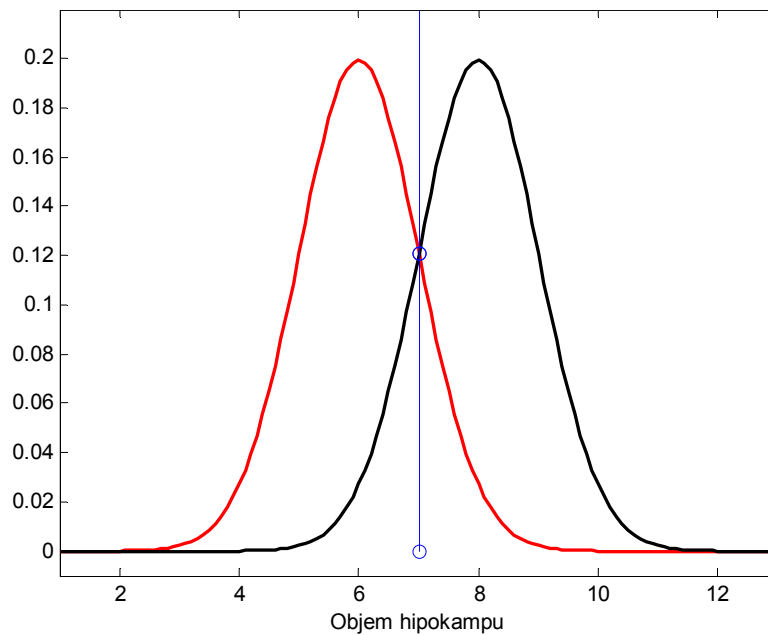
Výpočet podmíněné hustoty pravděpodobnosti výskytu obrazu ve třídě _____ :

$$\frac{\text{_____}}{\text{_____}} = \frac{\text{_____}}{\text{_____}}$$

Grafické znázornění podmíněných hustot pravděpodobnosti je znázorněno na Obrázku 5.

Výpočet celkové hustoty pravděpodobnosti:

Aposterioerní pravděpodobnosti: _____ a _____. Protože _____, nelze jednoznačně určit, do které třídy máme testovací subjekt zařadit. V takovém případě často klasifikační algoritmy náhodně zvolí jednu ze skupin.



Obrázek 5. Vizualizace hustoty pravděpodobnosti pacientů (znázorněna červeně) a kontrolních subjektů (znázorněna černě). Podmíněné hustoty pravděpodobnosti výskytu testovacího obrazu v jednotlivých třídách jsou znázorněny modře. Je patrné, že nelze rozhodnout, do jaké třídy máme testovací subjekt zařadit.

3. Klasifikace podle obou proměnných:

Spočteme aposteriorní pravděpodobnosti $P(C_1|X)$ a $P(C_2|X)$, přičemž $P(C_1) = 0.5$ a $P(C_2) = 0.5$, a zařadíme testovací subjekt do třídy s větší aposteriorní pravděpodobností.

Výpočet podmíněné hustoty pravděpodobnosti výskytu obrazu ve třídě C_1 :

$$f_{C_1}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$f_{C_1}(7) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(7-\mu)^2}{2\sigma^2}\right)$$

Výpočet podmíněné hustoty pravděpodobnosti výskytu obrazu ve třídě C_2 :

$$f_{C_2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$f_{C_2}(7) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(7-\mu)^2}{2\sigma^2}\right)$$

Grafické znázornění podmíněných hustot pravděpodobnosti je znázorněno na *Obrázku 6*.

Výpočet celkové hustoty pravděpodobnosti:

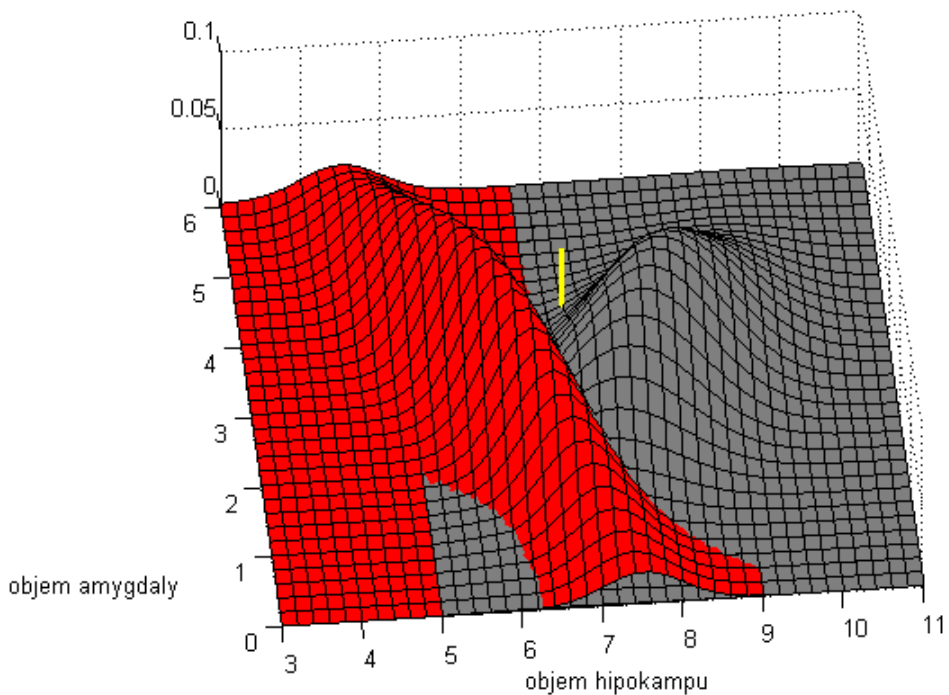
Aposteriorní pravděpodobnosti: _____ a _____ .

Protože _____ , zařadíme testovací subjekt do třídy kontrolních subjektů.

Výpočet hranice pomocí diskriminačních funkcí:

_____ → kritérium maximální aposteriorní pravděpodobnosti

Levá strana je rovna _____ a pravá strana rovna 1. Protože věrohodnostní poměr (na levé straně) je menší než výraz na pravé straně, subjekt zařadíme do třídy kontrolních subjektů.



Obrázek 6. Vizualizace hustoty pravděpodobnosti pacientů (znázorněna červenou plochou) a kontrolních subjektů (znázorněna šedou plochou). Podmíněná hustota pravděpodobnosti výskytu testovacího obrazu v jednotlivých třídách je znázorněna žlutě. Je patrné, že subjekt bude zařazen do třídy kontrolních subjektů.

Kritérium minimální pravděpodobnosti chybného rozhodnutí

Vydeme z výpočtu hranice pomocí diskriminačních funkcí (pro hranici je rozdíl diskriminačních funkcí roven 0).

$$\frac{1}{2} \left(\frac{1}{n_1} \sum_{i=1}^{n_1} x_i + \frac{1}{n_2} \sum_{i=1}^{n_2} x_i \right)$$

Můžeme vykrátit $\frac{1}{2}$, protože celková hustota pravděpodobnosti je stejná pro obě diskriminační funkce:

$\frac{1}{2} \left(\frac{1}{n_1} \sum_{i=1}^{n_1} x_i + \frac{1}{n_2} \sum_{i=1}^{n_2} x_i \right) \rightarrow$ kritérium minimální pravděpodobnosti chybného rozhodnutí

Výpočet pomocí dosazení do obecného vzorce pro výpočet věrohodnostního poměru, přičemž předpokladem je matice ztrátových funkcí ve tvaru $\begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$, potom získáváme:

$$\frac{1}{2} \left(\frac{1}{n_1} \sum_{i=1}^{n_1} x_i + \frac{1}{n_2} \sum_{i=1}^{n_2} x_i \right)$$

$$\frac{1}{2} \left(\frac{1}{n_1} \sum_{i=1}^{n_1} x_i + \frac{1}{n_2} \sum_{i=1}^{n_2} x_i \right)$$

$\frac{1}{2} \left(\frac{1}{n_1} \sum_{i=1}^{n_1} x_i + \frac{1}{n_2} \sum_{i=1}^{n_2} x_i \right) \rightarrow$ kritérium minimální pravděpodobnosti chybného rozhodnutí

Levá strana je rovna $\frac{1}{2} \left(\frac{1}{n_1} \sum_{i=1}^{n_1} x_i + \frac{1}{n_2} \sum_{i=1}^{n_2} x_i \right)$ a pravá strana rovna $\frac{1}{2} \left(\frac{1}{n_1} \sum_{i=1}^{n_1} x_i + \frac{1}{n_2} \sum_{i=1}^{n_2} x_i \right)$. Protože věrohodnostní poměr (na levé straně) je menší než výraz na pravé straně, subjekt zařadíme do třídy kontrolních subjektů.

Poznámka: Kdyby byly apriorní pravděpodobnosti jiné (v našem případě by se museli velmi lišit), např. $\frac{1}{2}$ $\frac{1}{2}$, v takovém případě by byl testovací subjekt zařazen do třídy pacientů.

Kritérium minimální střední ztráty

Pokud do výpočtu hranice pomocí diskriminačních funkcí zahrneme ztrátové funkce dané maticí ztrátových funkcí $\begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$, kde c_{11} je ztráta při klasifikaci kontrolního subjektu jako pacienta a c_{22} je ztráta při klasifikaci pacienta jako kontrolního subjektu etc. (přičemž vycházíme ze vztahu 2.23 na str. 17 ze skript), získáváme:

_____ → kritérium minimální střední ztráty

Levá strana je rovna _____ . Pravá strana je při různém nastavení vah rovna:

- A) _____ (tzn., více penalizují, pokud je pacient nesprávně zařazen do třídy kontrolních subjektů, než když je kontrolní subjekt nesprávně zařazen do třídy pacientů), pak pravá strana je rovna _____ a subjekt zařadím do třídy kontrolních subjektů. Museli bychom velmi penalizovat nesprávné zařazení pacienta do kontrolních subjektů (např. _____, aby byl testovaný subjekt vyhodnocen jako pacient.
- B) _____ (penalizují shodně nesprávné zařazení do třídy kontrolních subjektů i pacientů – kritérium minimální pravděpodobnosti chybného rozhodnutí), pak pravá strana je rovna _____ a subjekt zařadím do třídy kontrolních subjektů.
- C) _____ (tzn., více penalizují, pokud je kontrolní subjekt nesprávně zařazen do třídy pacientů, než když je pacient nesprávně zařazen do třídy kontrolních subjektů), pak pravá strana je rovna _____ a subjekt zařadím do třídy kontrolních subjektů.

Kritérium maximální pravděpodobnosti

Předpokladem je rovnoměrné zastoupení tříd, tzn. _____, a nulové ztráty při správném rozhodnutí, tzn. _____, pak získáváme po dosazení do obecného vzorce pro výpočet věrohodnostního poměru:

_____ → kritérium maximální pravděpodobnosti

Levá strana je rovna _____ . Pravá strana je při různém nastavení vah rovna:

- A) _____ (tzn., více penalizují, pokud je pacient nesprávně zařazen do třídy kontrolních subjektů, než když je kontrolní subjekt nesprávně zařazen do třídy pacientů), pak pravá strana je rovna – _____ a subjekt zařadím do třídy kontrolních subjektů.
- B) _____ (penalizují shodně nesprávné zařazení do třídy kontrolních subjektů i pacientů), pak pravá strana je rovna – _____ a subjekt zařadím do třídy kontrolních subjektů.
- C) _____ (tzn., více penalizují, pokud je kontrolní subjekt nesprávně zařazen do třídy pacientů, než když je pacient nesprávně zařazen do třídy kontrolních subjektů), pak pravá strana je rovna – _____ a subjekt zařadím do třídy kontrolních subjektů.