

Odhad pohlaví a populační afinity Základy

Mgr. Mikoláš Jurda, Ph.D.

Diskriminační analýza

Závislost jedné kvalitativní proměnné (konečný počet a priori známých tříd) na několika kvantitativních proměnných

Skupina objektů se známou vlastností
(**trénovací množina**)

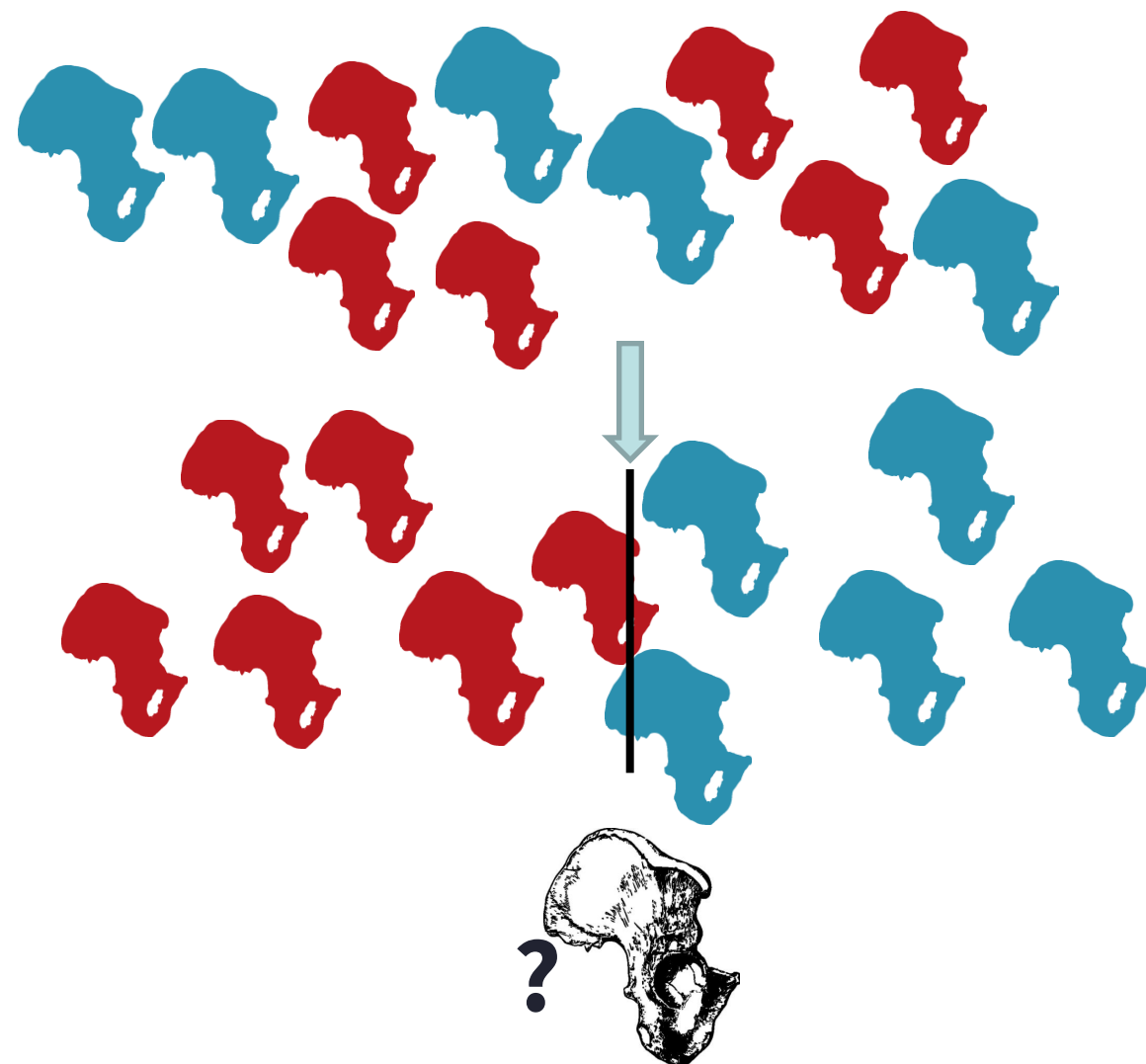


Rozhodovací pravidlo (**diskriminační rovnice**)
– lineární kombinace proměnných, která
nejlépe diskriminuje mezi známými skupinami

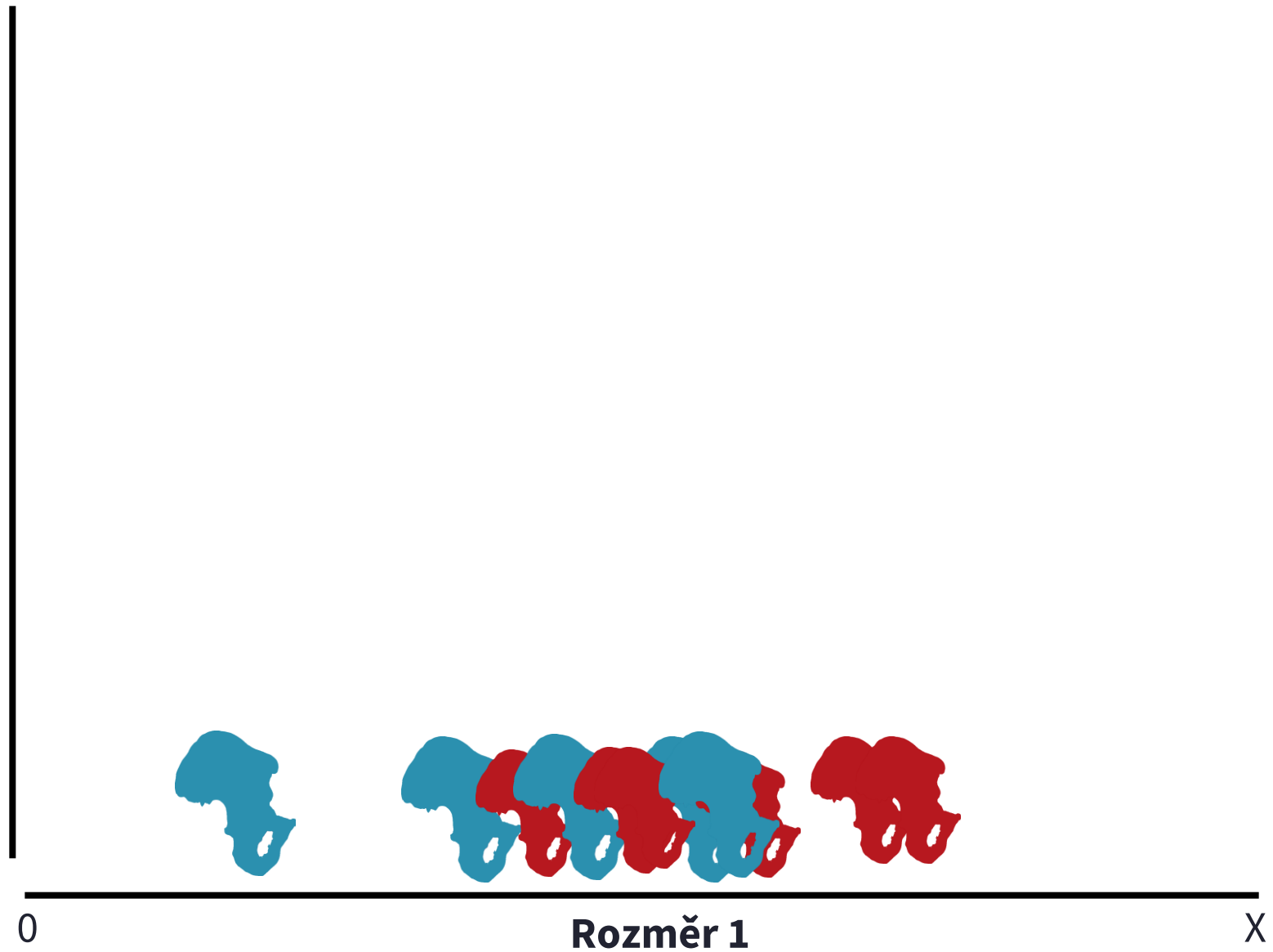


Klasifikace neznámých případů

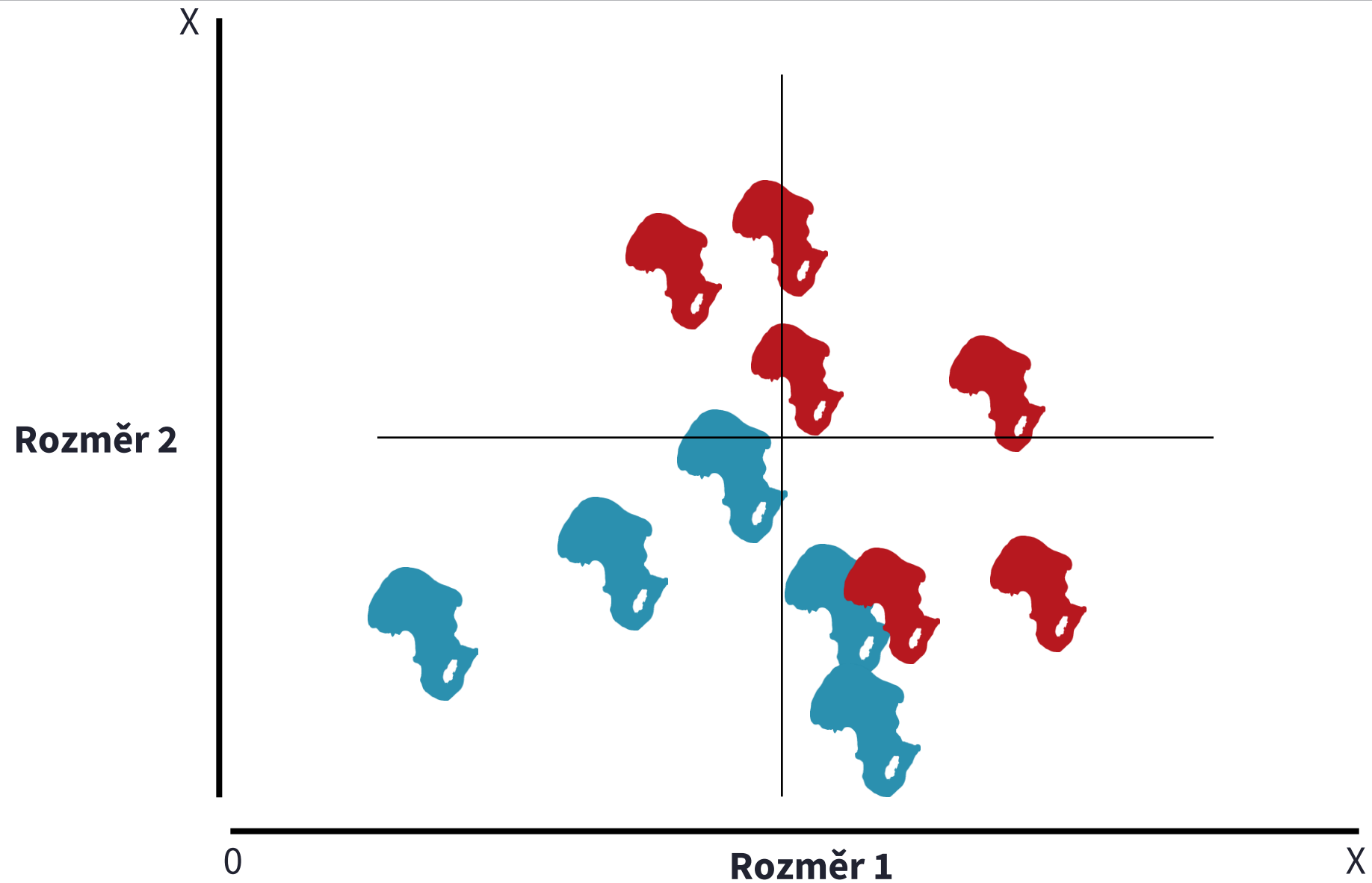
Správnost můžeme očekávat pouze
pokud je ze stejné populace



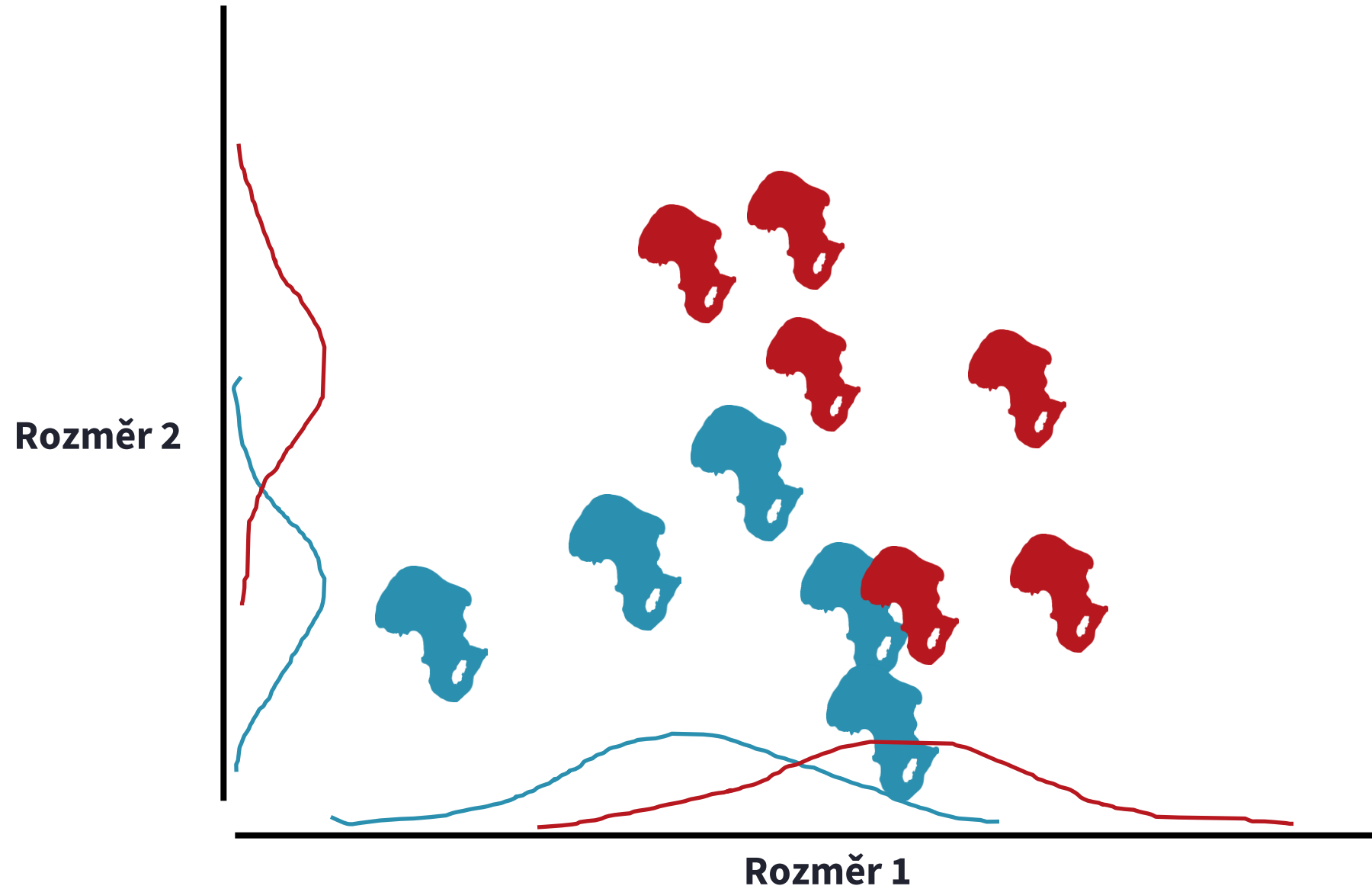
Diskriminační analýza



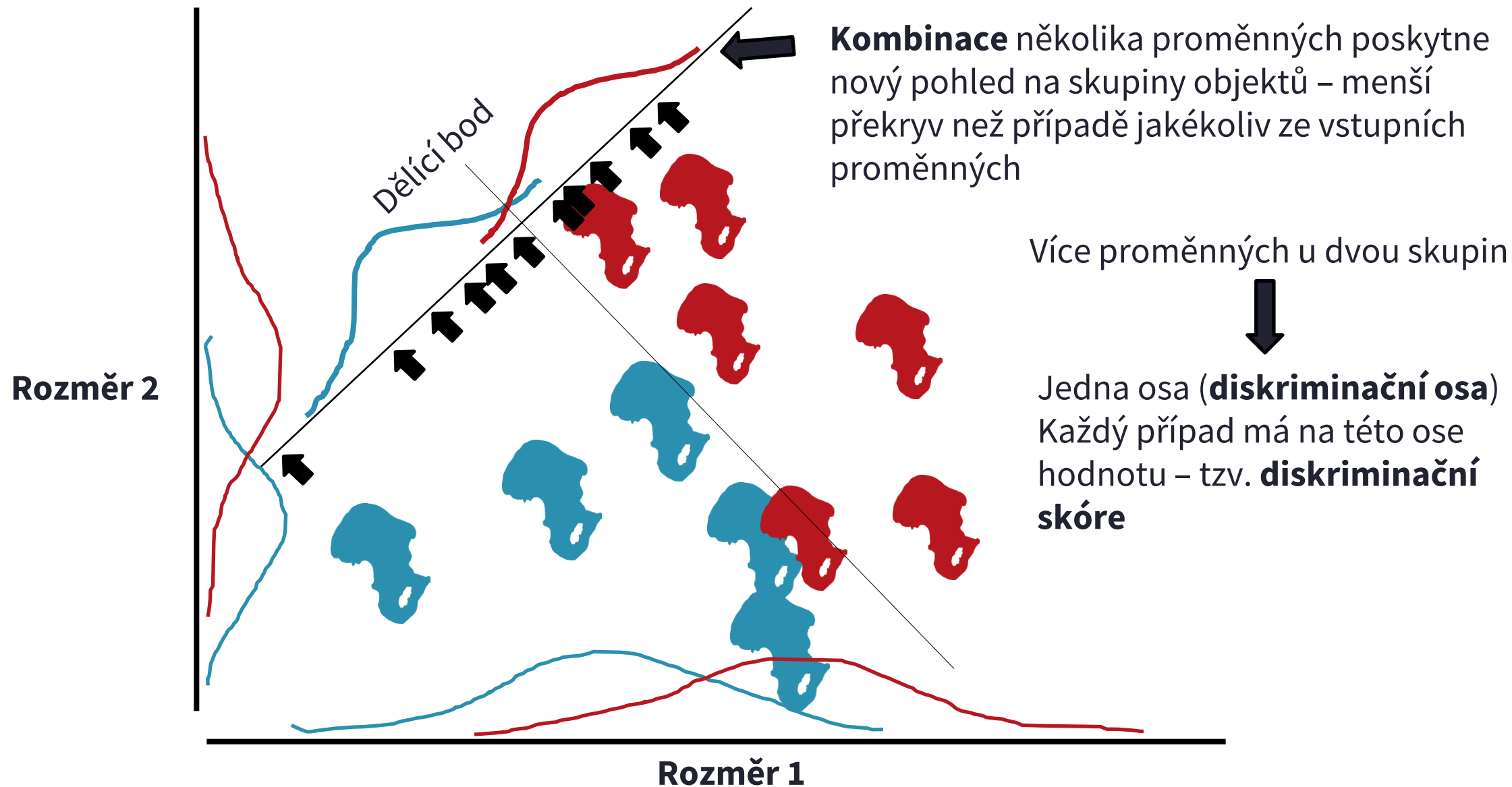
Diskriminační analýza



Diskriminační analýza



Diskriminační analýza



Diskriminační analýza

Lineární kombinace původních proměnných -
klasifikační rovnice

Alternativní zápisy

$$S1 = 1,270 \times \text{míra1} + 0,232 \times \text{míra2} + 103,502$$

konstanta

suma proměnných
násobených koeficienty

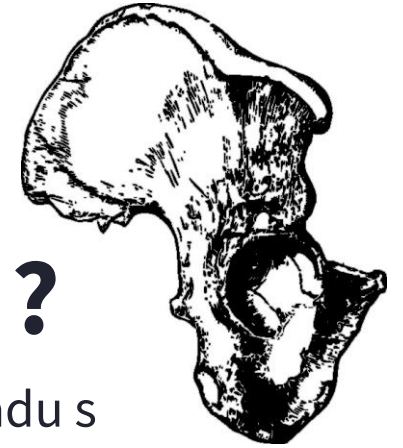
Rozměr1
Rozměr2

1,270
0,232
103,502

Koeficienty vypovídají
o důležitosti
jednotlivých
proměnných

Porovnání skóre
neznámého případu s
dělicím bodem

Výsledek



Více rovnic – určuje například výše výsledku (Novotný
et al. 2003, nebo rovnice pro odlišení více skupin

Diskriminační analýza

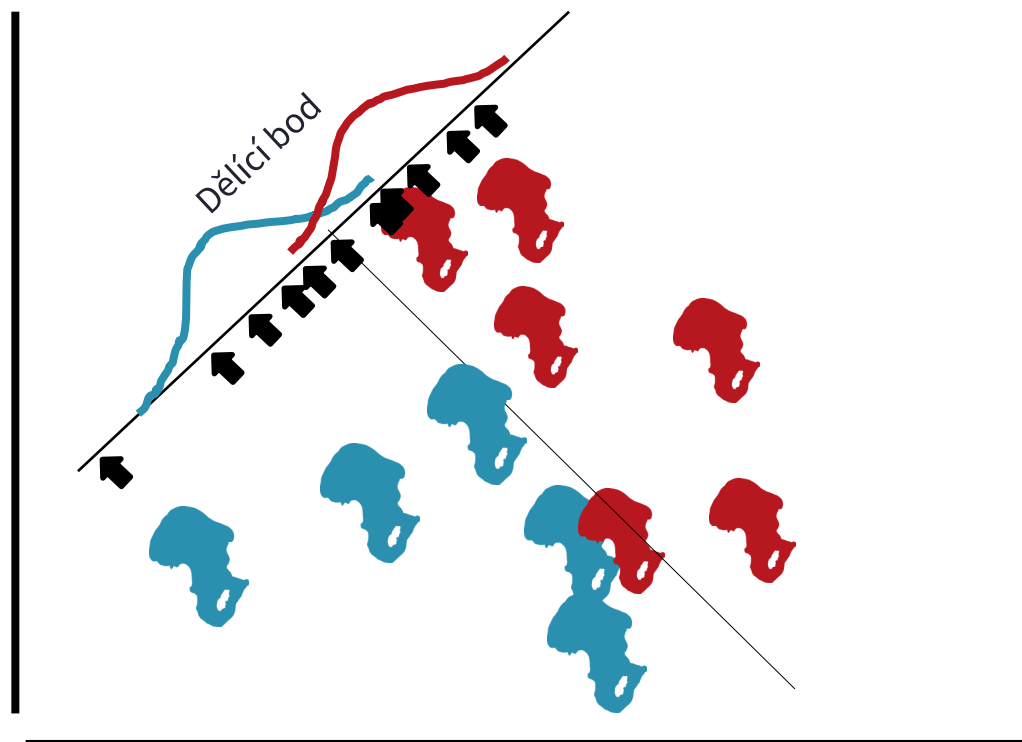
Hodnota diskriminačního skóre

vs.

dělicí bod

vs.

rozložení trénovací množiny

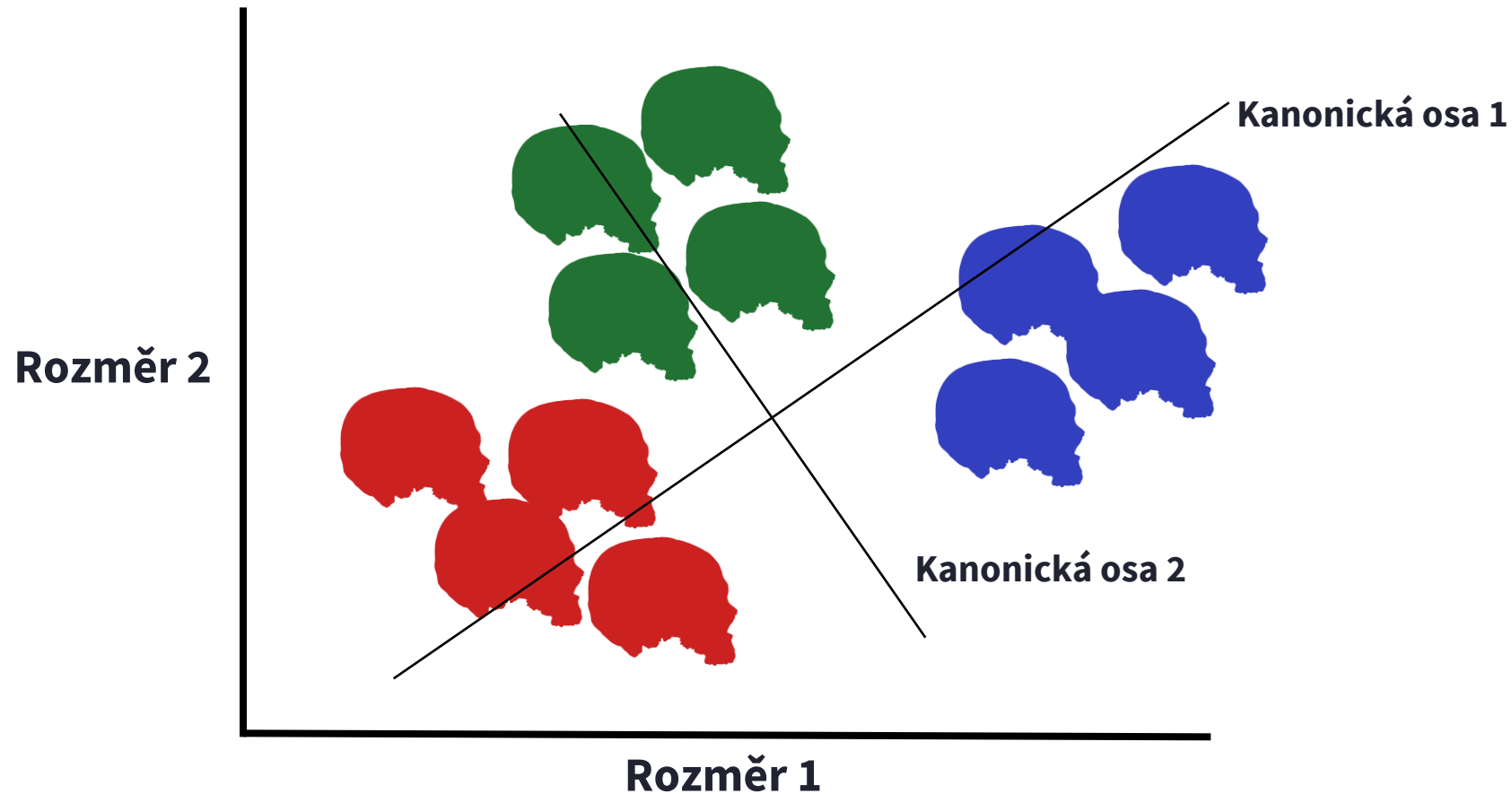


Posterioční pravděpodobnost (*posterior probability*) – relativní vzdálenost případu (skóre) k centroidu každé skupiny (v součtu jsou jedna, hodnota tedy závisí také na množství skupin)

Analýza kanonických proměnných

Projekce **vícerozměrného prostoru** maximalizující diskriminaci mezi dvěma a více skupinami. Možné používat například ke zobrazení největších rozdílů mezi skupinami.

Rozšíření diskriminační analýzy na více než dvě skupiny.



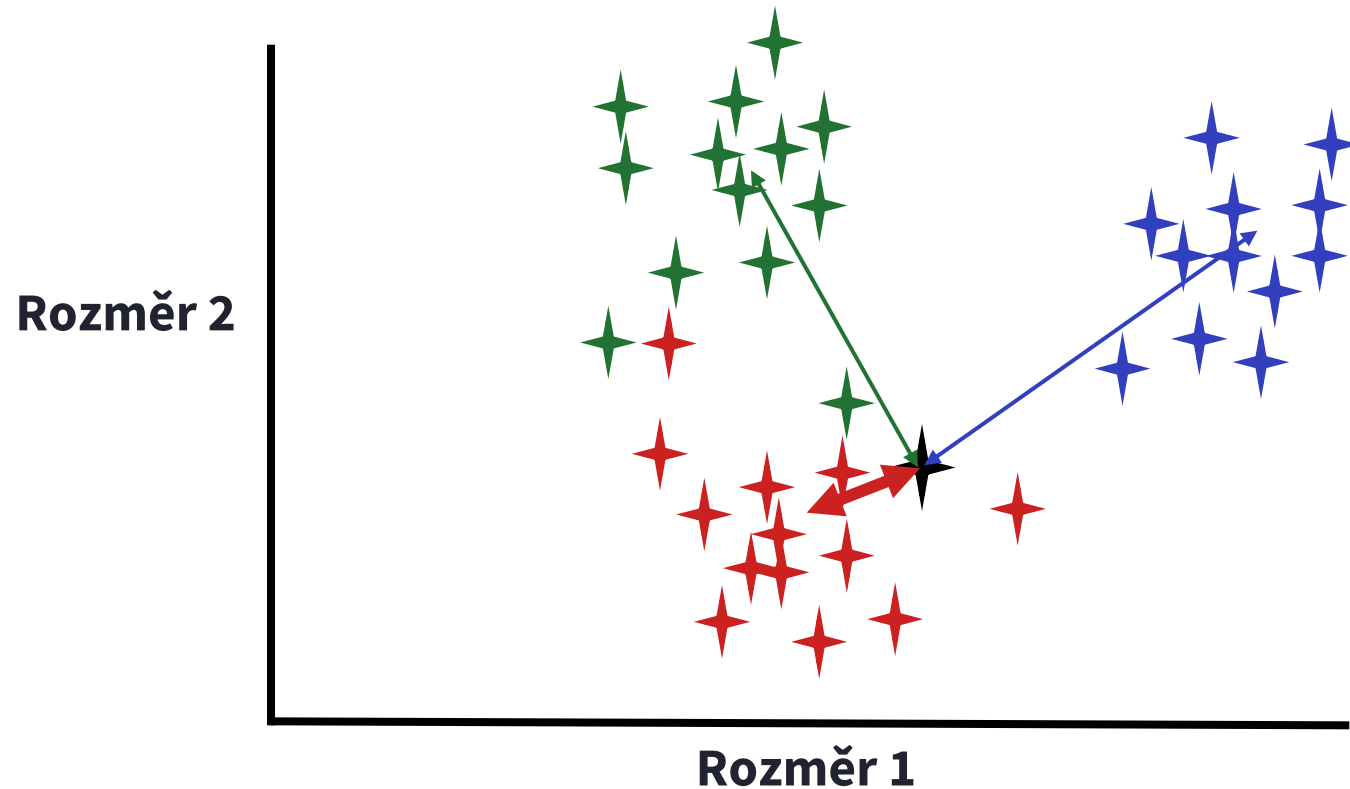
Více proměnných u
N skupin



N-1 kanonických os
(seřazeny podle
klesající důležitosti)
Každý případ má
skóre na každé ose

Analýza kanonických proměnných

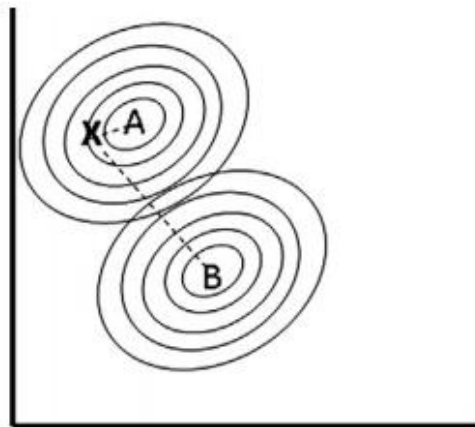
Klasifikace na základě **vzdálenosti klasifikovaného případu k centroidu jednotlivých skupin** známých případů ve vícerozměrném prostoru (počet rozměrů je dán počtem proměnných)



Zobrazit ale můžete jen první dvě nebo tři osy, ale z pohledu ostatních os to může vypadat jinak

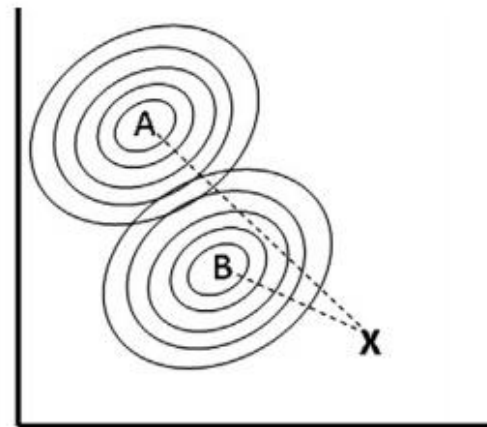
Analýza kanonických proměnných

Typikalita (*typicality probability*) – míry typičnosti daného případu pro každou ze skupin. Vysoká hodnota znamená, že daný případ je typický zástupce dané skupiny a naopak



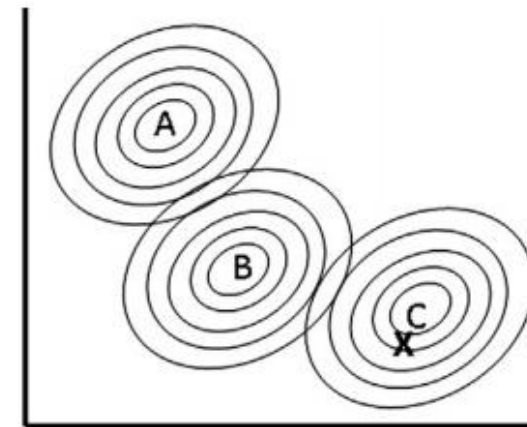
Posterior probability of X to
A: 0.99
B: 0.01

Typicality probability of X to
A: 0.800
B: 0.001



Posterior probability of X to
A: 0.001
B: 0.999

Typicality probability of X to
A: 0.000
B: 0.001



X is very unusual for group A or B, and
could belong to a third group, C

Účinnost klasifikace

Resubstitute (pozorovaná účinnost, *observed accuracy rate*)

Klasifikujeme stejné případy, na kterých bylo kritérium vytvořeno

Klasifikace testovacího souboru

Klasifikujeme jiné případy ze stejné nebo jiné populace (často třeba druhou polovinu dostupného vzorku)

Křížové ověření (leave-one-out crossvalidation)

Ze souboru n objektů vybereme $n - 1$ objektů, které použijeme jako trénovací soubor, z něhož odvodíme klasifikační kritérium. Toto pak aplikujeme na jeden vypuštěný případ.

		Odhadnutá hodnota	
		Skupina 1	Skupina 2
Pozorovaná hodnota	Skupina 1	12	1
	Skupina 2	2	11

Specifická a senzitivita

$$\text{senzitivita} = \frac{\text{počet skutečně pozitivních}}{\text{počet skutečně pozitivních} + \text{počet falešně negativních}}$$

Kolik případů test správně odhalí.

Například kolik mužů správně určí jako muže.

$$\text{specifická} = \frac{\text{počet skutečně negativních}}{\text{počet skutečně negativních} + \text{počet falešně pozitivních}}$$

Kolik případů test odhalí nesprávně.

Například kolik žen určí jako muže.

		Odhadnutá hodnota	
		Muž	Žena
Pozorovaná hodnota	Muž	30	0
	Žena	15	15

FORDISC 3.0 (Jantz & Ousley 2005)

Část KS: lebka a postkraniální skelet

Metody: lineární míry, analýza kanonických proměnných

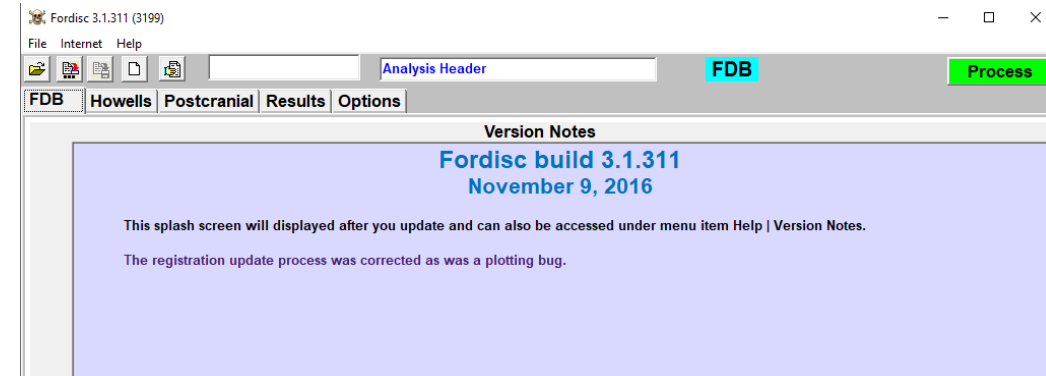
Nástroje: Dotykové měřidlo, posuvné měřidlo, počítač

Populace: Howells, Terry, Hamann-Todd (16. století př. k. – 20. století) a FDB forenzní databáze (+ od 1986)

Celková přenost: Brazilci různého původu (19. a 20. století) 50 % (FDB) a 44,5 % (Howwellova databáze; Urbanová & Jurda 2014); vybrané skupiny z Howellovy databáze (-1600-20. století) pod 40 % (Elliott & Collard 2012); novověcí Španělé (Madridská sbírka celkově 53,68 %

Poloautomaticky fungující počítačový software, s poměrně jasnými pravidly

Pohlaví je odhadováno spolu s etnickým původem, neoddělitelně!!!



FORDISC 3.0 (Jantz & Ousley 2005)

FDB databáze

- obsahuje (mimo jiné) osteometrická data recentních forenzních případů (od roku 1986) a pozitivně identifikovaných původních obyvatel Spojených Států
- součástí je rozsáhlá dokumentace obsahující údaje o místu narození, věku, zraněních, váze, výšce



White m/f – Euroameričané, někteří narozeni v Evropě

Black m/f – Afroameričané, celé území Spojených Států

Hispanic m/f – etnická příslušnost určena „podle kontextu“, většina z Nového Mexika

American Indian m/f – for. případy z jihozápadu US + pozitivně identifikované případy z 19. století

Japanese m/f – region Kanto, Japonsko

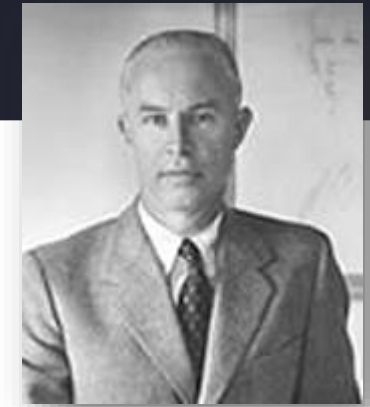
Vietnamese m – oběti rudých Kmérů z Kambodži

Chinese m – Honk-kong

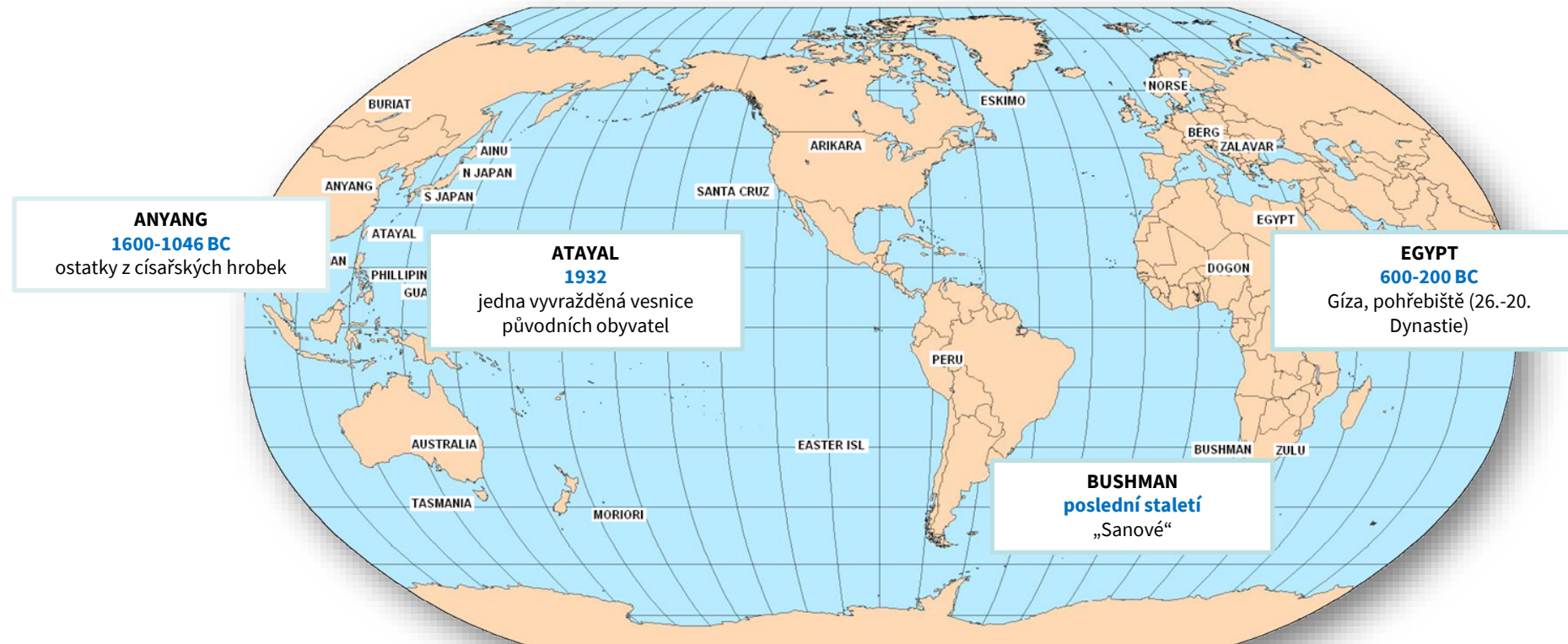
Guatemalan m – recentní případy

FORDISC 3.0 (Jantz & Ousley 2005)

- obsahuje osteometrická data více než 2500 tisíc lebek, až 70 rozměrů
- + *Terry a Hamann –Todd collections* (150 *American Blacks*, 174 *American Whites*) obsahující lebky z 19. a 20. století.
- **heterogenní** jak do způsobu definování populací, tak časově
- pohlaví až na výjimky **určeno Howellsem inspekcí!!!**



William **White** Howells



FORDISC 3.0 (Jantz & Ousley 2005)

**Fordisc vždy poskytne nějaký výsledek!!!
Nesmí se používat otrocky!**

1. Zadejte hodnoty a relevantní populace – pokud zadáte atypickou hodnotu, program Vás upozorní.

Výběr populací?

Jaké populace
můžeme očekávat?

Pro moderní nálezy je
vhodná FDB, pro historické
Howellsova databáze)

Měli by být určování pouze ti
jedinci, kteří náleží do
referenčních populací

Pokud můžeme spolehlivě určit pohlaví,
například podle postkraniálu, pak
můžeme výběr omezit pouze na dané
pohlaví

Zároveň sledujte početnost zbývajících populací. **Nejméně početná populace by měla mít třikrát více jedinců než je proměnných.**

V případě, že toto nesplňujete, využijte step-wise analýzu anebo odstraňte proměnné.

FORDISC 3.0 (Jantz & Ousley 2005)

Multigroup Classification of Current Case

Group	Classified into	Distance from	Probabilities			
			Posterior	Typ F	Typ Chi	Typ R
WF	**WF**	5.5	0.990	0.915	0.905	0.851 (38/249)
AF		15.9	0.005	0.497	0.144	0.226 (25/31)
HF		17.7	0.002	0.230	0.089	0.123 (51/57)
BF		18.4	0.002	0.161	0.072	0.143 (67/77)
JF		19.3	0.001	0.160	0.056	0.063 (60/63)

Current Case is closest to WFs

Mahalanobisova D2

Posteriovní probabilita

Různé typikality (standardně Typ F)

2. Nevěnujte pozornost pouze určené populaci, ale také **dalším poskytnutým hodnotám!!!**
3. Analýzu opakujte bez populací, **kteří mají PP menší než 0,1 a TP menší než 0,01**. Pokud tyto kritéria splní některá další populace, znovu analýzu opakujte bez ní.

Výsledky lze přijmout pouze **pokud PP přesáhne 0,5 a TP 0,01 nebo 0,05** (manuál programu)

3D-ID

Reference: SLICE, Dennis E a Ann H. ROSS, 2009. 3D-ID: geometric morphometric classification of crania for forensic scientists. Version 1.0

Část KS: lebka

Metody: souřadnice bodů, Prokrustovská superpozice (možno omezit na tvarové proměnné), analýza kanonických proměnných

Nástroje: digitizér, digitální modely, PC

Přesnost: Brazilci (+1917-1937) 55 % (rozdělení do tří skupin), Evropané 83 % (Urbanová et al. 2014)

Diagnóza probíhá podobně jako ve Fordisc

Proměnných by mělo být více než jedinců v referenčních populacích, každá souřadnice je jedna proměnná!!!

Díky využití souřadnic analýza pracuje v větším množství informací!!!

Výhodou jsou geograficky vymezené populace a zařazení hned několika moderních Evropských sbírek

3D-ID

1. Zadejte souřadnice vybraných bodů – vyvarujte se záměny levých a pravých bodů
2. V záložce *Options* vyberte relevantní populace
3. Spusťte analýzu příkazem *Process* v první záložce

Interpretujte všechny číselné výsledky a analýzu případně opakujte

```
Summary...
=====
                                D2          Posterior  Typicality
African - female (3):          141,6543    0,0000     0,0188
African - male (4):            146,9295    0,0000     0,0037
African_American - female (97): 118,6565    0,0000     0,0038
African_American - male (122):  109,3626    0,0002     0,0138
African_Brazilian - female (7):  123,2596    0,0000     0,0151
African_Brazilian - male (7):   112,3218    0,0003     0,0516
Brazilian - female (9):         127,8168    0,0000     0,0054
Brazilian - male (31):           95,8352    0,0823     0,0959
Circumcaribbean - female (1):  191,3152    0,0000     0,0701
Circumcaribbean - male (3):    161,0573    0,0000     0,0022
East_Asian - male (19):         108,1043    0,0007     0,0299
European_American - female (62): 100,6479    0,0098     0,0470
European_American - male (98):   89,7973    0,9013     0,1443 <====
```

Walrath et al. 2014

WALRATH, D.E., P. TURNER a J. BRŮŽEK, 2004. Reliability test of the visual assessment of cranial traits for sex determination. *American Journal of Physical Anthropology* 125, 132–137.

Součást KS: lebka

Metoda: vizuální, 10 znaků

Populace: Severní Amerika (Inuité, Aljaška) – 500-900 AD

Celková senzitivita: N/A

Poměrně dobře hodnotitelné a popsané znaky.

Vizuální zhodnocení
znaků

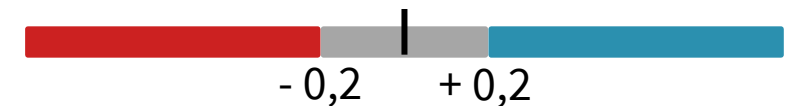


Výpočet **indexu sexualizace** z
přítomných znaků

$$IS = \frac{\Sigma (score \times weight)}{\Sigma weight}$$



Hodnota určuje pohlaví



Pokud IS vyjde $\pm 0,2$, hodnotí se jako indiferentní

Publikace: BRŮŽEK, Jaroslav, Frédéric SANTOS, Bruno DUTAILLY, Pascal MURAIL a Eugenia CUNHA, 2017. Validation and reliability of the sex estimation of the human os coxae using freely available DSP2 software for bioarchaeology and forensic anthropology: BRŮŽEK et al. American Journal of Physical Anthropology. 164(2), 440–449. ISSN 00029483.
doi:10.1002/ajpa.23282

Metoda: metrická, 4-10 rozměrů, Fisherova lineární diskriminační analýza, pohlaví je určeno, pokud pp překoná 0,95

Vybavení: posuvné měřidlo, PC

Populace: Olivierova sbírka, Tamagnini sbírka, Spitalfields, Garmusova sbírka, Dartova sbírka – Zulu, Dartova sbírka – Soto, Dartova sbírka – „Afrikaner“, Hamann-Toddova sbírka – Euroameričané, Hamann-Toddova sbírka – Afroameričané, Terryho sbírka – Euroameričané, Chang-Mai sbírka

Celková senzitivita: původní soubor - v závislosti na kombinaci proměnných 98,67 - 99,65 % u 41,49 - 90,98 případů; Maxwelllova sbírka, Simonova sbírka – 50,86 – 99,02 podle použitých proměnných