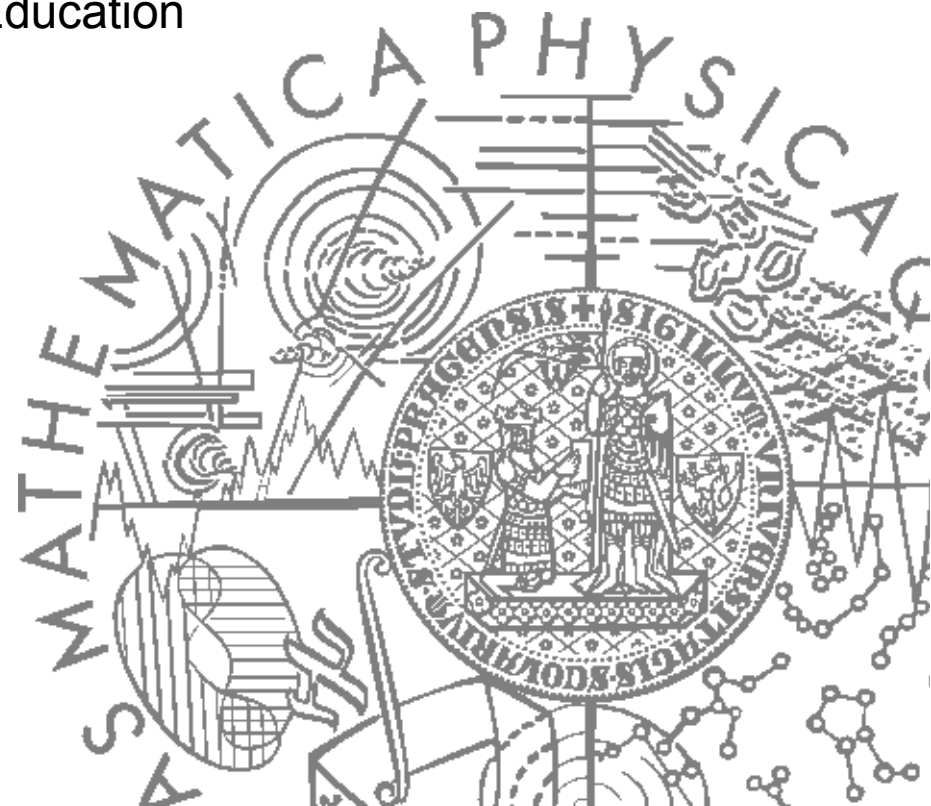


# Software pro 2D geometrickou morfometrii

Václav Krajíček

Vaclav.Krajicek@mff.cuni.cz

Department of Software and Computer Science Education  
Faculty of Mathematics and Physics  
Charles University



# Osnova

---

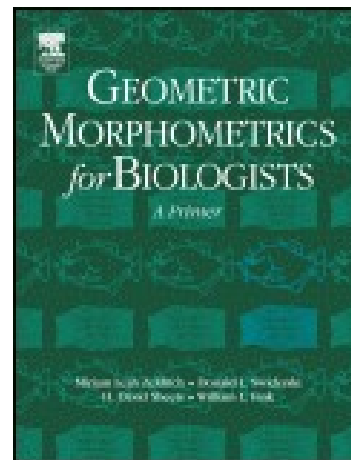


- Úvod
- Morfometrická analýza
- Hypotéza
- Sběr dat
- Měření
- Zpracování
- Vzdálenost
- Velikost
- Tvar
- Závěr

# Úvod

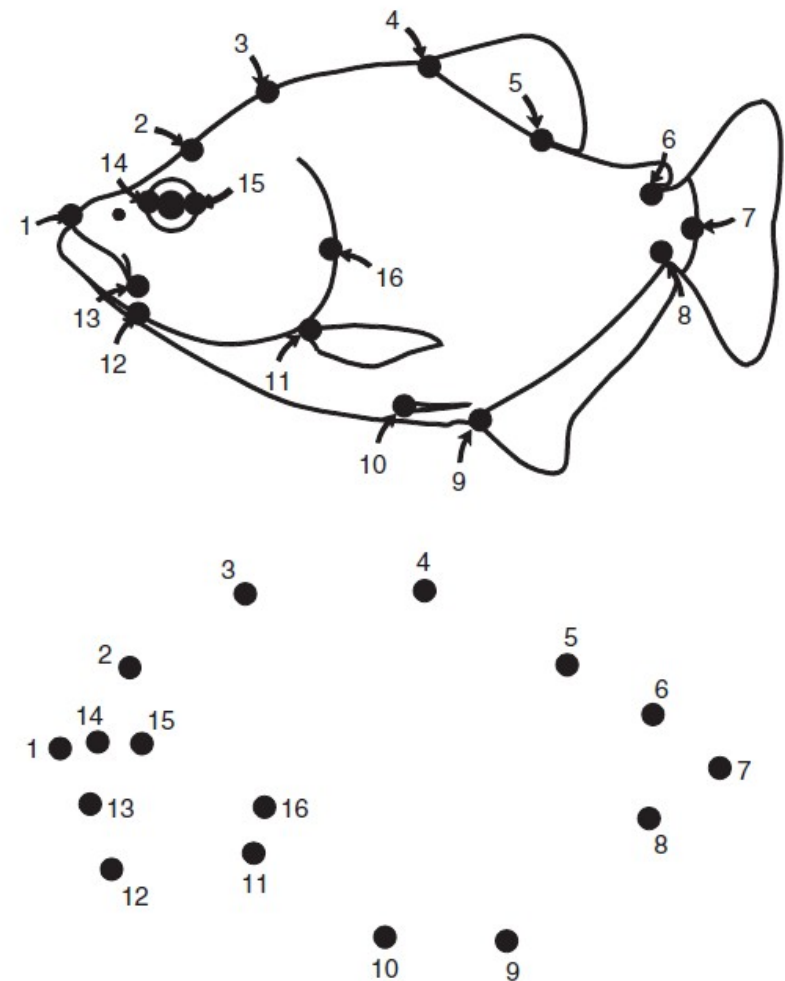
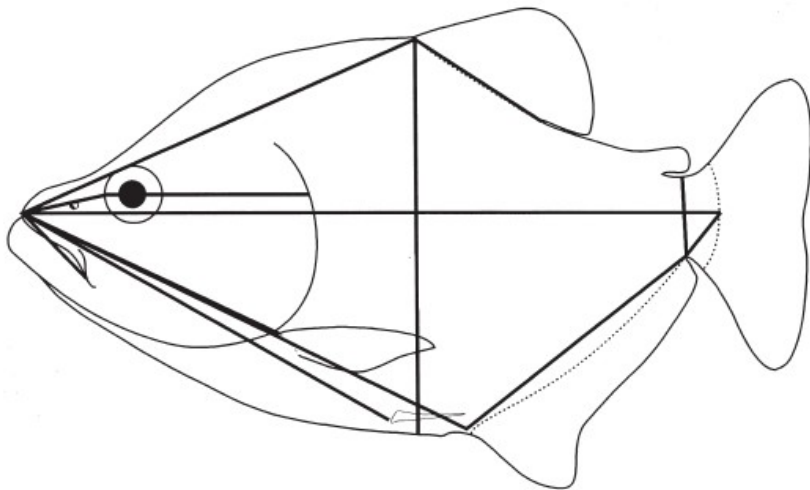


- Andrea Cardini: GMM Workshop York 2008
- Zelditch, Swiderski, Sheets, Fink: GMM: A Primer



# Morfometrická analýza

- „Tradiční“ morfometrie
- Geometrická morfometrie



# Hypotéza



- „Musíme vědět co chceme zjistit“
- Nenechat se unést dostupností software/metod
  - Je zvolená metoda vhodná pro můj problém?
  - Má náš problém odpovídající předpoklady požadované metodou?
- Subjektem hypotézy je celá populace, ale pracovat typicky můžu jen se vzorkem.
  - Velikost
  - Reprezentativnost

# Sběr dat



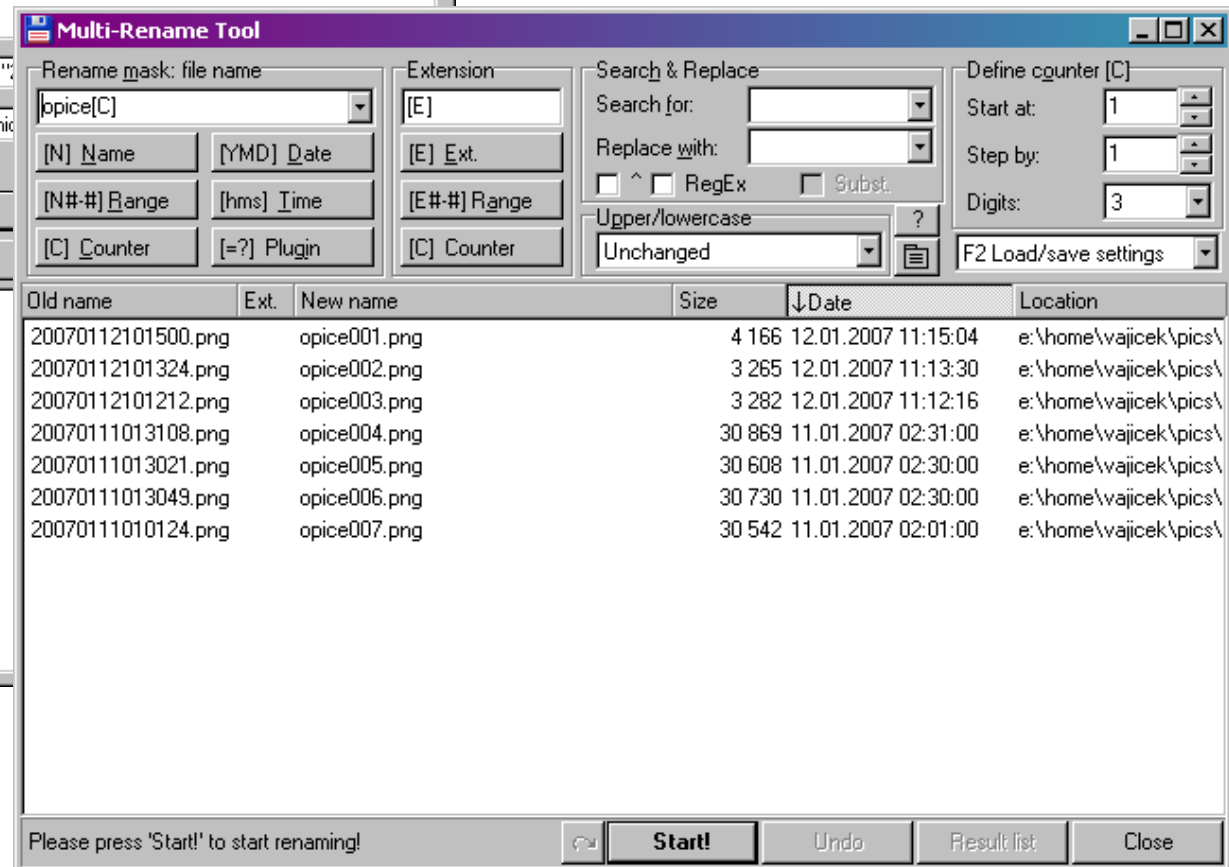
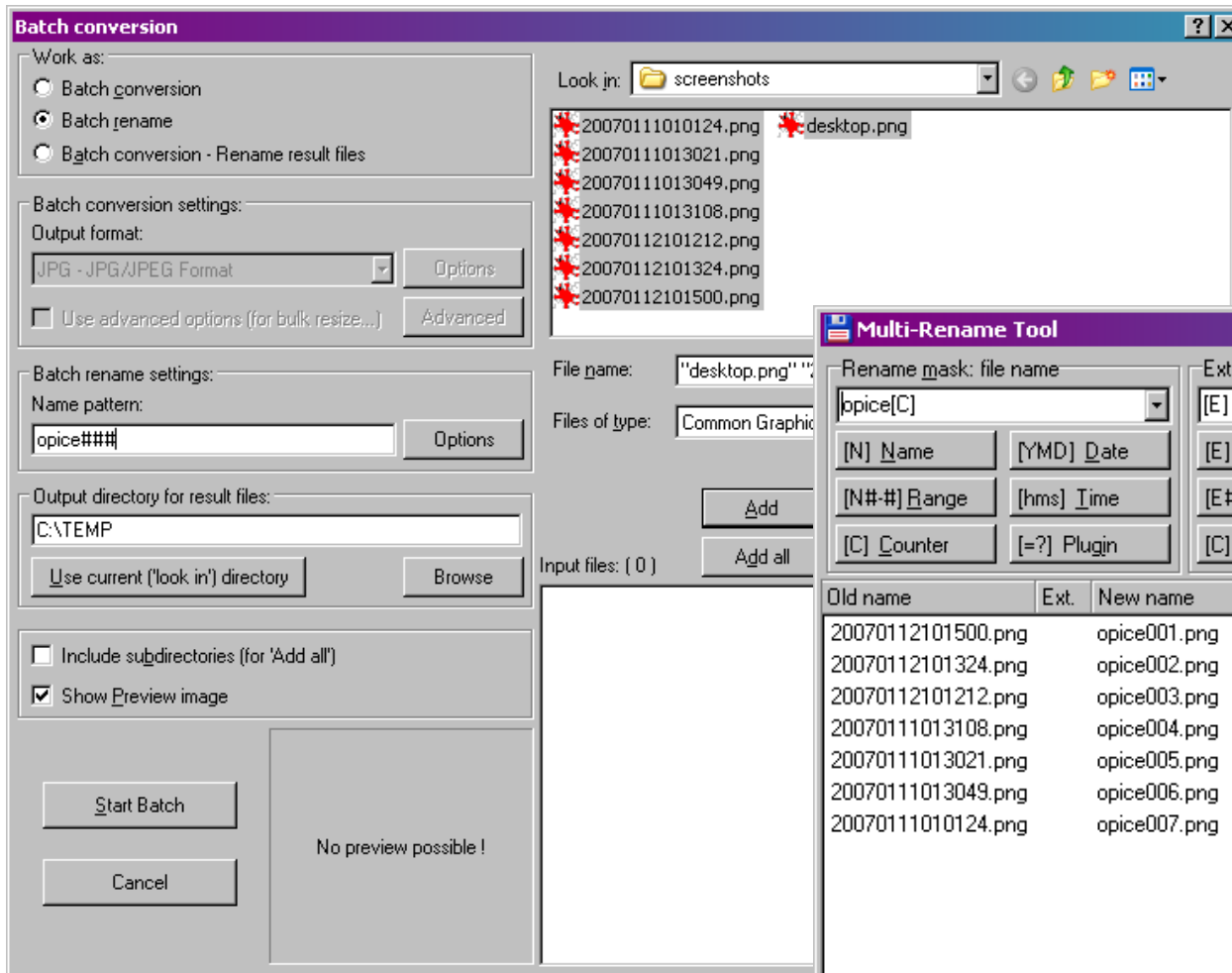
- Digitální fotografie/obraz
- Dokumentační/forenzní fotografie
- Důkladná příprava
  - možnost zavlečení systematické chyby
- Základní pravidla
  - Stativ, vzdálenost
  - Osvětlení, přímé, z více zdrojů, redukce stínu
  - Kontrastní pozadí
  - Značky, měřítko
  - Větší vzdálenost + zoom na max → rovnoběžné paprsky

# Digitální fotografie



- Obrazové formáty se kterými pracuje software
  - jpeg, gif, png, DICOM
  - Rozlišení podle potřeby
    - Moc velké → pomalá práce programů při načítání
    - Moc malé → ztráta detailu
- Pojmenování souborů
  - Z dig. fotoaparátu – IMG0131.jpg, IMG0135.jpg
  - Hromadné přejmenování
    - Total commander, Irfan
  - Identifikace obrázků podle obsahu

# Hromadné přejmenování





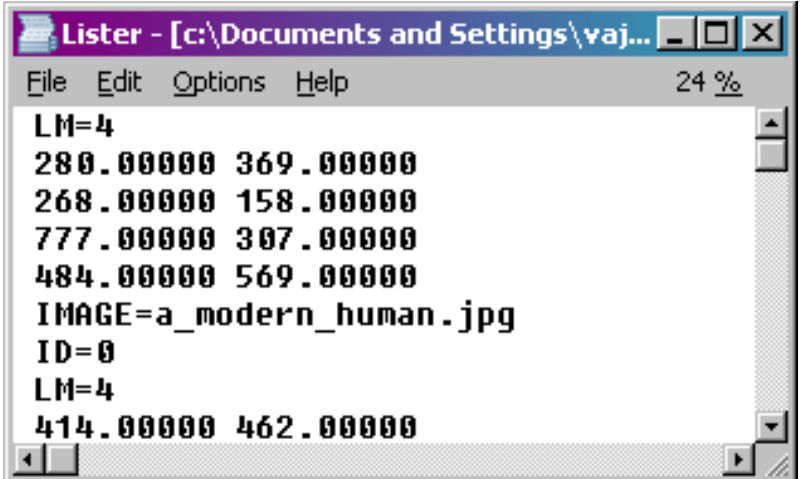
# Sběr dat

- **tps** programy
  - Malé jednoduché, nestabilní
- Formáty souborů
  - Soubor je jednotka informace na disku počítače
    - Textové soubory – editovatelné v notepadu
    - Binární soubory – speciální nástroje, rychlejší práce programů, nečitelné člověka



```

Lister - [e:\home\vajicek\pics\screenshots\... 5 %
File Edit Options Help
[PNG#####IHDR#####i#####ÇÃp'##=IIDA
0'''6##H#Eb'##±ã»=óù>Á###!###Ù###(ââ, D2#c#Èu
#~#0_ wu0[##$«0#:#ëõê###`sîîî.7%#çw~ç09###
»Úòzð0\¼###okz»¿ÿ%³½{:#ù¿##Çûx0/##íppêç?=<
óð, #èP#t##L#:#z0###äC#ô5NPò##|Ë_ú/?ÿó#à;¾
á###%çÇ###RDÉ# # ('W5#0í#0#êL#ê#JpÉ##È+¾ð0
þÅGÿ#CÉ# # ýòAÉãt##L#:#S##aà#ou³Níãryxx''Ú
#÷ðé0þòíí7<|æ0ÿë#ùííÛQ2#####JîUMãt##L#:#S
#_ùð÷=ýµ#îíÿô#ÿà_ ù#/#ú##,xó³ûâ#/#i;¾###ü#/#¿
  
```



```

Lister - [c:\Documents and Settings\vaj... 24 %
File Edit Options Help
LM=4
280.000000 369.000000
268.000000 158.000000
777.000000 307.000000
484.000000 569.000000
IMAGE=a_modern_human.jpg
ID=0
LM=4
414.000000 462.000000
  
```

# Formáty souborů

.TPS

```
LM=4
280.00000 369.00000
268.00000 158.00000
777.00000 307.00000
484.00000 569.00000
IMAGE=a_modern_human.jpg
ID=0
LM=4
414.00000 462.00000
239.00000 190.00000
792.00000 455.00000
600.00000 553.00000
IMAGE=baboon.jpg
ID=1
```

.NTS

```
" links file
1 4 2 0
1 2
2 3
3 4
1 4
```

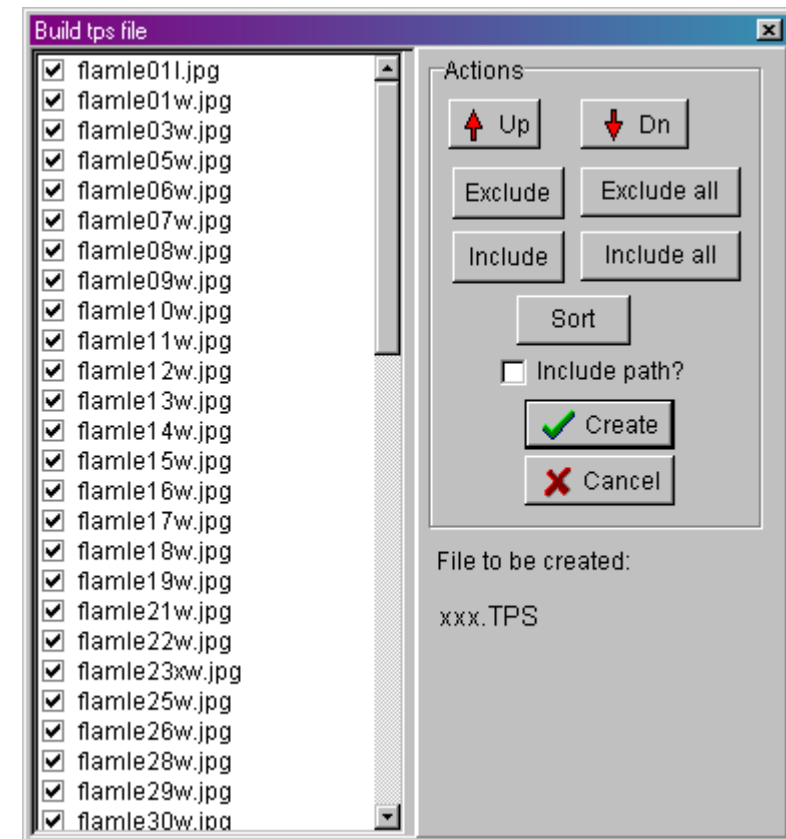
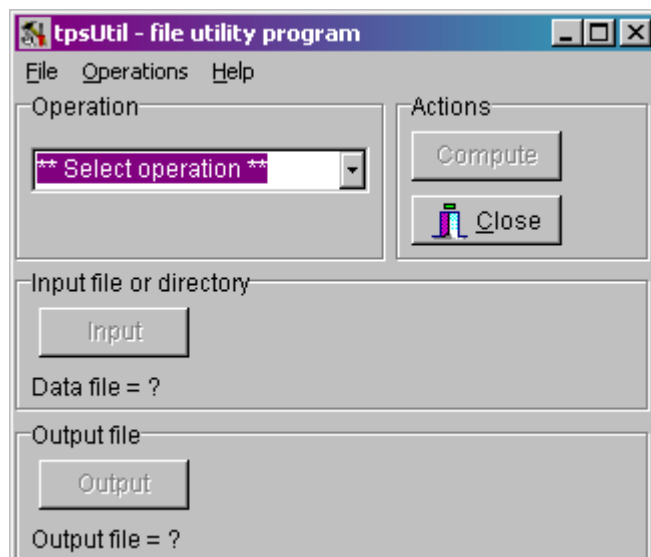
.CSV

```
11;12;13
21;22;23
31;32;33
```

- nápověda např. tpsSmall

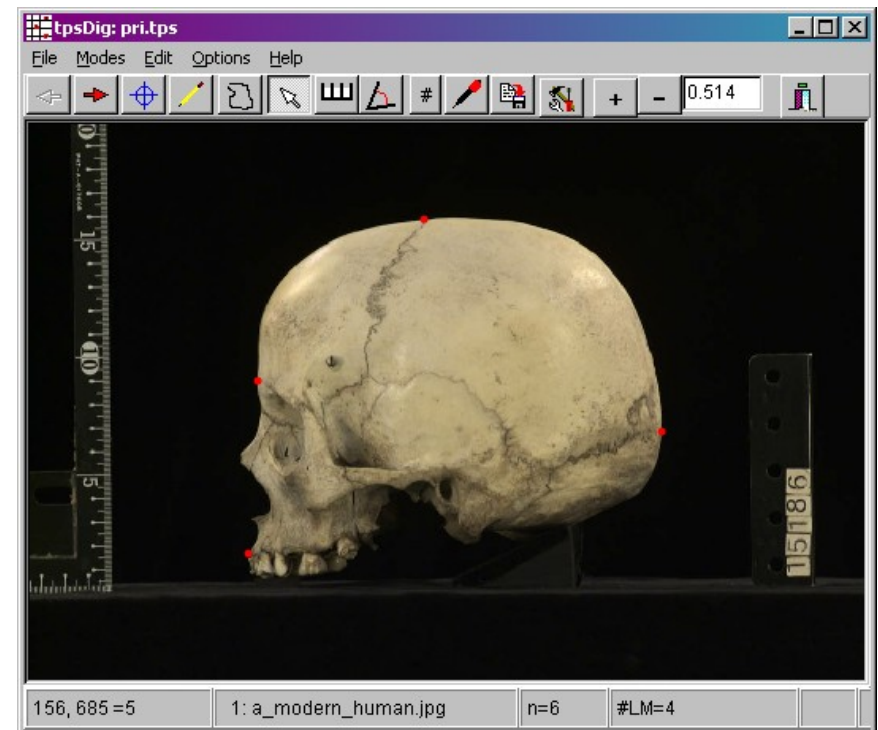
# Příprava měření

- Vytvoření kolekce/souboru/množiny
- tpsUtil
  - Funkce, které se jinde nevěšly
  - „build tps file from images“



# Měření

- tpsDig, tpsDig2
- Načtení vytvořeného souboru
- Kalibrace
  - Může platit pro všechny / pro každý zvlášť
- „Naklikat“ landmarky
  - Na každém obrázku
    - Stejný počet
    - Stejně pořadí
- Uložit jako .TPS/.NTS



# Zpracování dat

- Máme čísla a chceme z nich něco počítat
- Software
  - Obecný
    - R, Matlab, S++, Excel
    - Učebnice
  - Speciální
    - PAST, tpsRegr, tpsRelw, tpsSmall
    - Návod
- Načítání dat
  - .TPS → PAST
  - PAST → Excel
  - Excel → .CSV

# Příklad 1

- *Marmota himalayanus* / *Marmota flaviventris*



# Analýza vzdáleností 1

- **Hypotéza:** Druh *himalayanus* ma větší čelist než *flaviventris*
- **Měření:** Dvojice bodů na množině, def. **vzdálenosti**
  - data/2D means
  - tpsUtil, tpsDig
  - uložit soubor TPS
- **Zpracování:** Načíst soubor v programu PAST
  - Dvě možnosti:
    - Výpočet přímo v PAST (geomet → Distance from landmarks)
    - Překopírovat do Excelu (vzorec pro vzdálenost bodů)

$$d(A, B) = \sqrt{(Ax - Bx)^2 + (Ay - By)^2}$$

# Analýza vzdáleností 2

- **Zpracování:**

- Dvě skupiny vzdáleností
- T-test

$$T = \frac{\bar{X} - \bar{Y} - \delta}{\sqrt{(n-1)S_x^2 + (m-1)S_y^2}} \sqrt{\frac{nm(n+m-2)}{n+m}}$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left[ \sum_{i=1}^n X_i^2 - \frac{1}{n} \left( \sum_{i=1}^n X_i \right)^2 \right]$$

- Excel, PAST

from Wikipedia: The Free Encyclopedia.

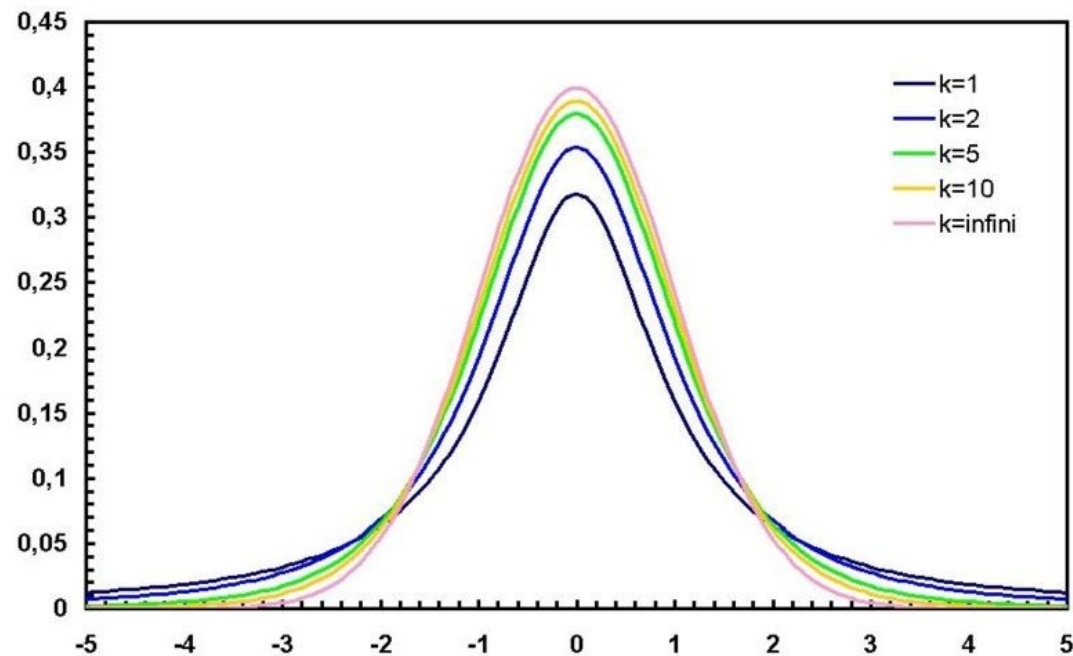
- **Kontrola:**

- Výpočet v různých programech (chyba v postupu)
- Větší množiny, permutační testy (chyba ve vzorku)



# Analýza vzdáleností 3

- **Závěr:** výsledkem t-testu je
  - t hodnota, interpretace tabulky (hodnoty) t rozdělení
  - p hodnota, „pravděpodobnost, že dvě množiny stejné“

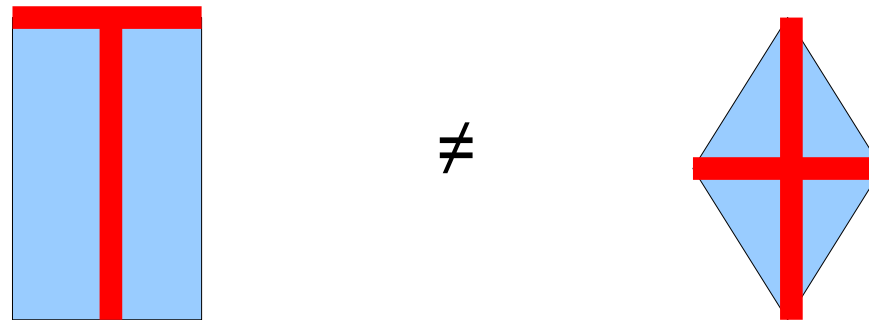


# Permutační testy

- V případě malého vzorku z velké populace
- Zvýšení jistoty správného výsledku
- Postup:
  - Dvě porovnávané množiny měření spojit
  - Zamíchat
  - Rozdělit na nové dvě množiny
  - Vypočítat t-test
  - Poznamenat si jestli vyšel hůř nebo lépe
  - Opakovat mnohokrát (1000x)
- Vyzkoušet si
  - `other\software\other programs\rtest2andrea.xls`

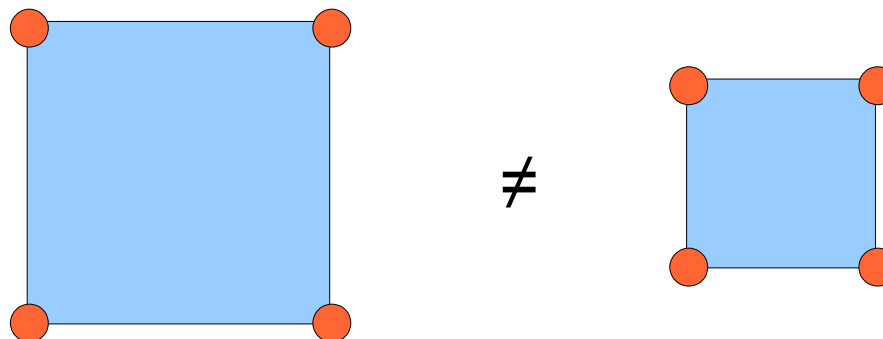
# Tvar a velikost

- Vzdálenosti ani poměry vzdáleností nestačí



- Dva objekty mohou být identické tvarem, ale mohou mít jinou velikost.

- Nedají se porovnávat na základě polohy landmarků



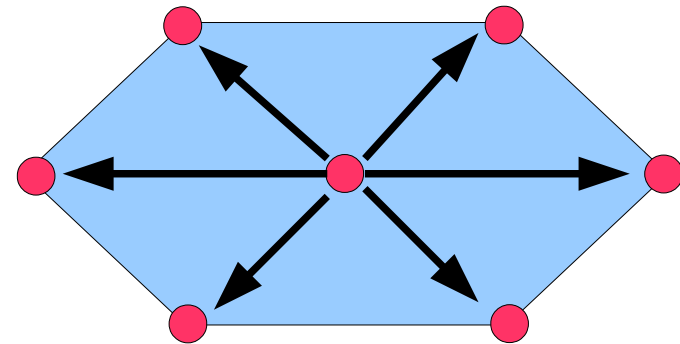
# Analýza velikost

- Těžiště

$$\bar{c}_k = \frac{1}{N} \sum_{i=1}^N c_{k,i}$$

- Centroidová velikost (centroid size)

$$\bar{s}_k = \sqrt{\sum_{i=1}^N (c_{k,i} - \bar{c}_k)^2}$$



# Příklad 2

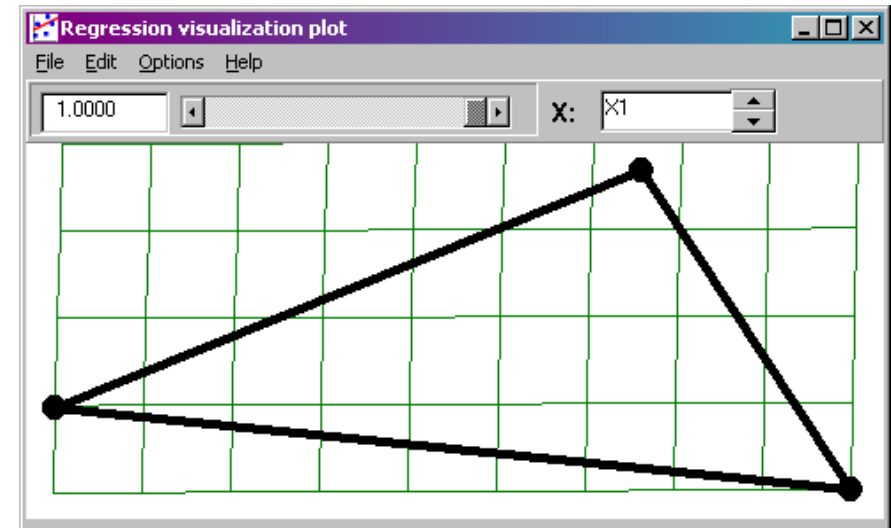
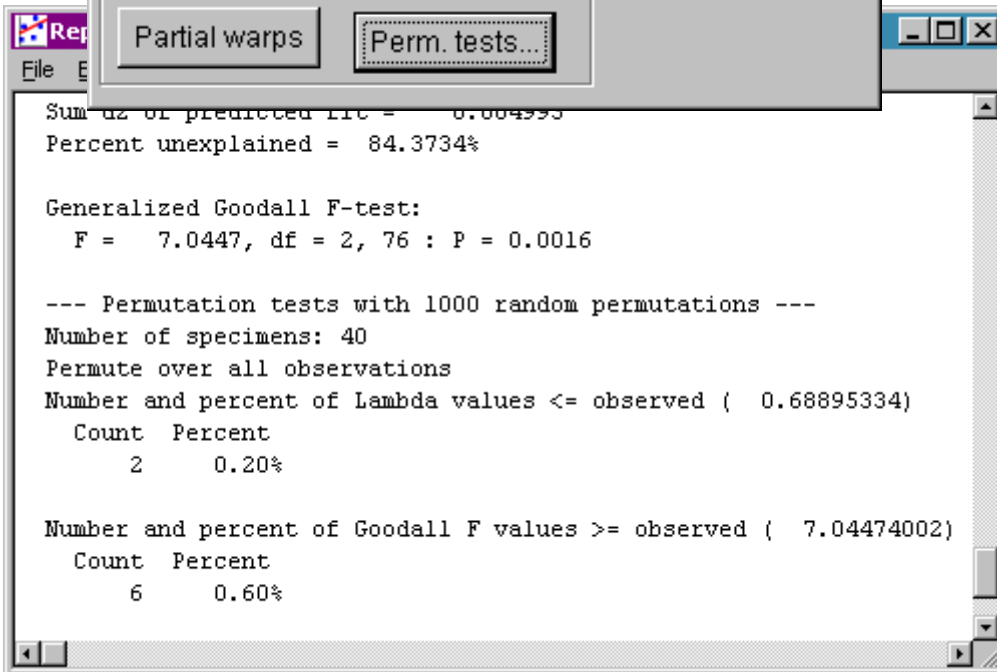
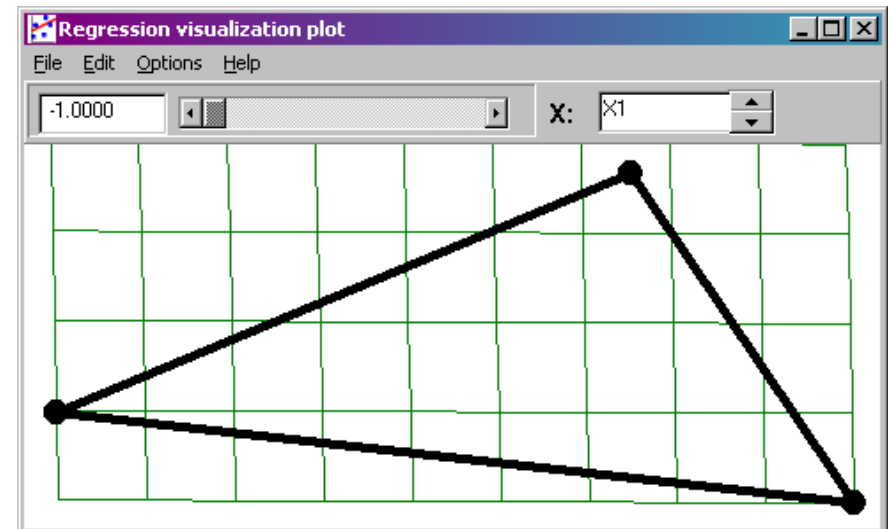
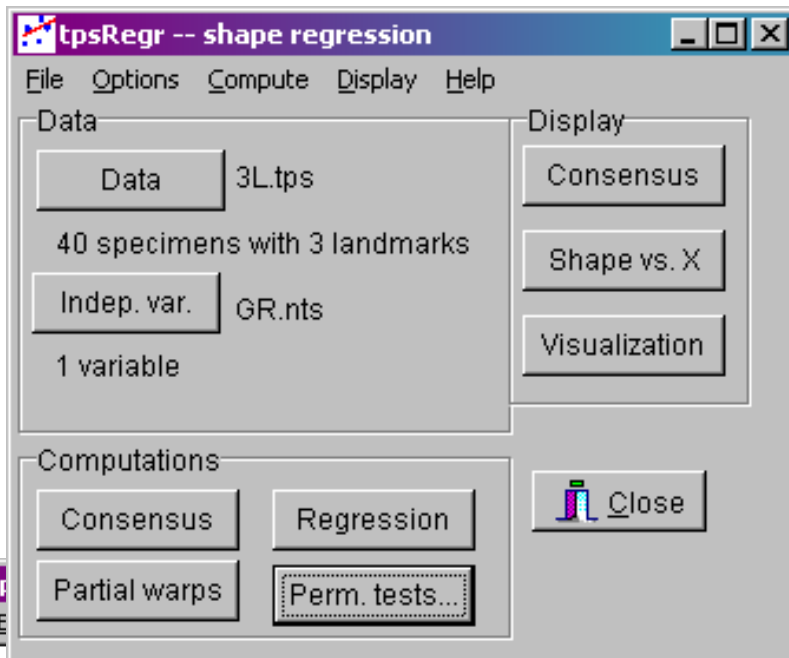
- *Marmota himalayanus* a *flaviventris* vzhledem k centroidové velikosti
- Vztah velikosti a pohlaví u *Marmota flaviventris*
- Matematický nástroj: **lineární regrese**

$$Y = f(X, \beta)$$

lineární vztah mezi závislými a nezávislými proměnnými

- Software: **tpsRegr**
  - tpsDig – 3 landmarky na jedince
  - (links.nts – tabulka 3x2, seznam spojníc landmarků)
  - gr.nts – tabulka počet jedinců x 1, nezávislá proměnná

# Regrese velikosti a druhu



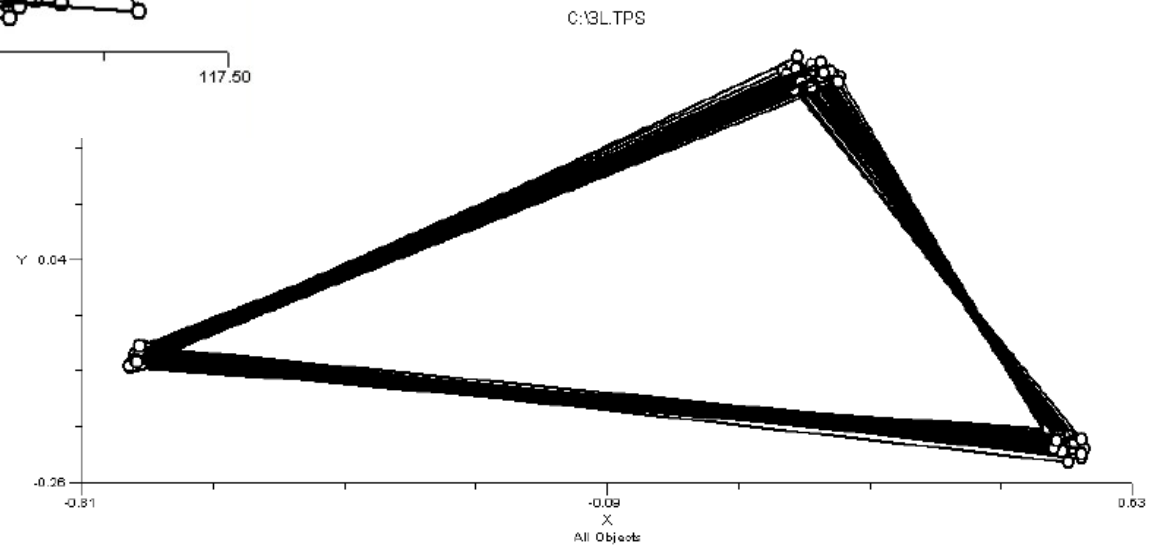
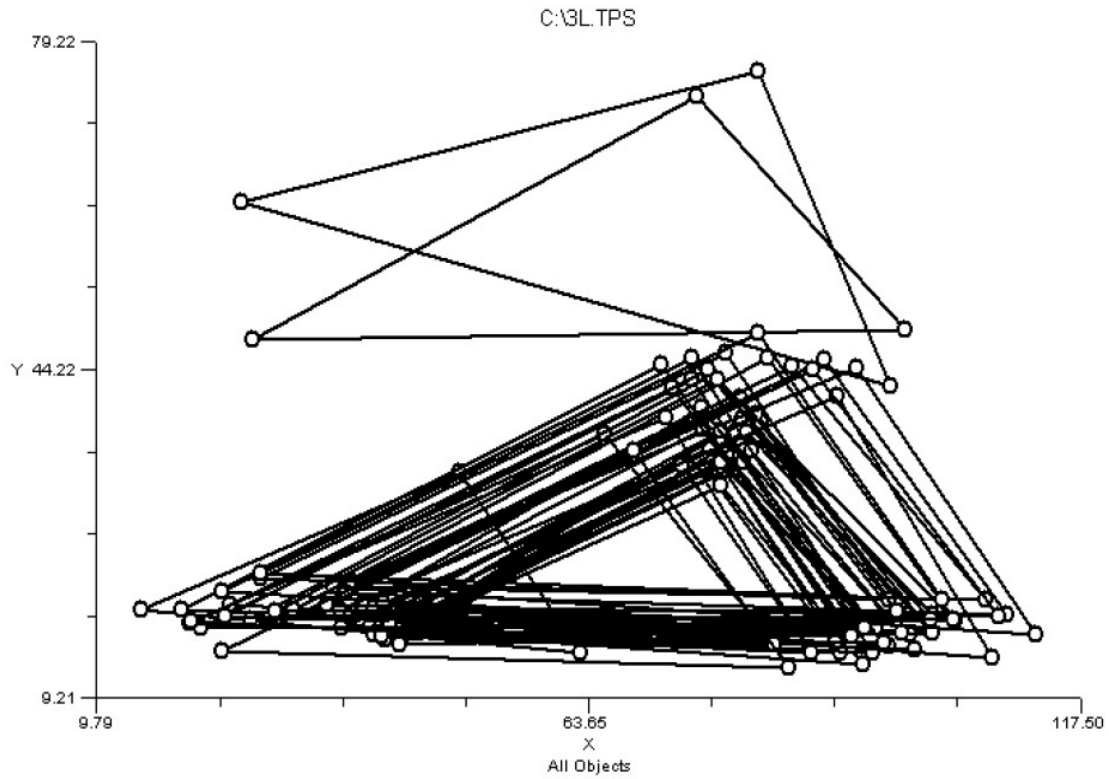
# Popis tvaru

- Odstranění velikosti, polohy a otočení
- Prokrustovská analýza (GPA, rigidní registrace)
  - Matematicky: minimalizace chyby

$$E = \sum_{i=1}^k \sum_{j=1}^n \text{dist}(l_{i,j} - \bar{l}_i)$$

- Algoritmicky:
  - Posuň všechny tak, aby měly těžiště v počátku
  - Naškáluj všechny, aby centroidová velikost rovna prvnímu
  - Otáčeť každý tak, aby měl nejmenší vzdálenost od středního tvaru
  - Vypočti nový střední tvar a opakuj pokud se hodně liší od předchozího
- Jiné typy normalizace

# Ukázka





# Deformace

- Protože máme korespondenci několika bodů ve dvou obrázcích můžeme se pokusit dopočítat korespondenci libovolných bodů
- Pro účely vizualizace
- Matematický prostředek: TPS interpolační funkce (Thin plate splines)

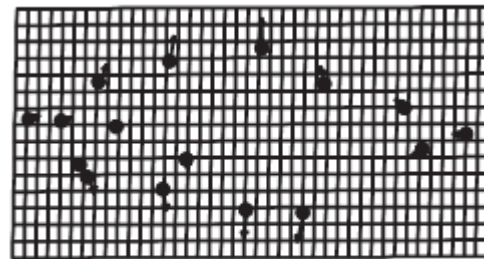
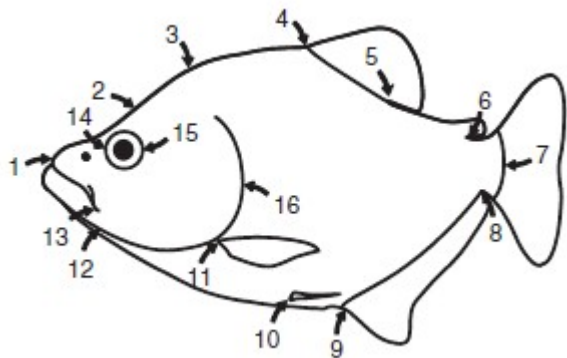
$$f_{tps}(z, \alpha) = f_{tps}(z, d, c) = z \cdot d + \sum_{i=1}^K \phi(\|z - x_i\|) \cdot c_i$$

- $d$  – parametry afinní složky,  $c_i$  – parametry deformace
- $\Phi$  – radiální báze

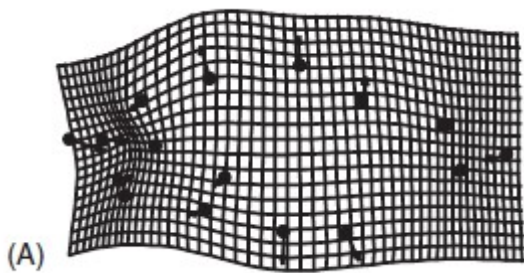
- Software: tpsRelw

# Ukázka

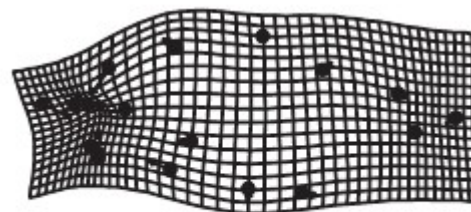
- Vizualizace mřížkou



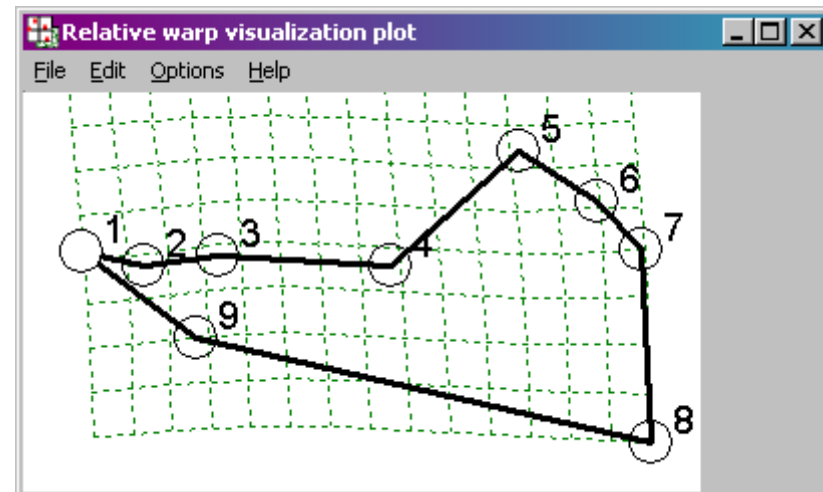
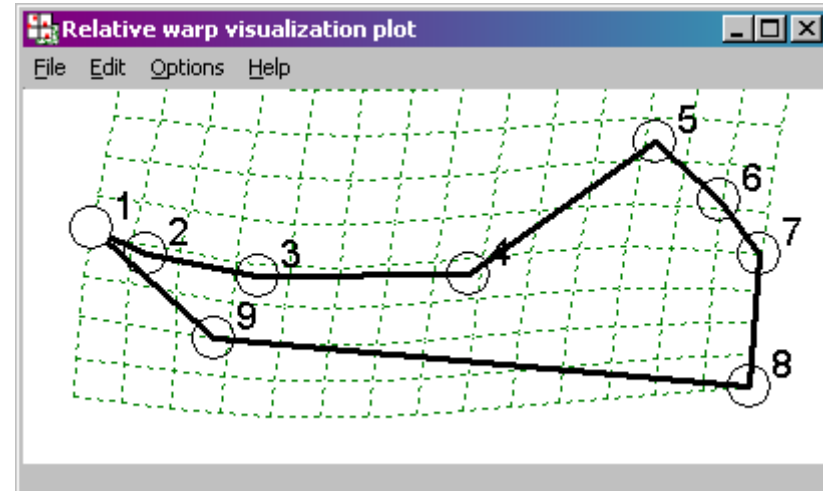
(B)



(A)



(C)



# Porovnávání tvarů

- Dvě množiny landmarků, dvou jedinců
  - Zarovnané pomocí GPA
  - Jak hodně jsou si podobné?
- Eukleidovská vzdálenost, Procrustes shape distance

$$d = \sqrt{(u_1 - x_1)^2 + (v_1 - y_1)^2 + (u_2 - x_2)^2 + (v_2 - y_2)^2 + \dots}$$

- Tato míra méně odpovídá skutečnosti čím je větší
  - Zakřivení prostoru tvarů
  - Pracujeme v rovinné aproximaci (jako papírová mapa)
  - Eukleidovská vzdálenost je přiblížením skutečné vzdálenosti v prostoru tvarů

# Příklad 3

- Určení příbuznosti na základě podobnosti tvarů
- Shluková analýza
  - V každém kroku sloučím jedince kteří jsou si v prostoru tvarů nejbliže
- Postup:
  - `data\2D distances\`
  - TpsUtil, tpsDig
  - TpsRegr, concensus, file → Save → Aligned specimens
  - Nebo přímo v PAST (transform → Procrustes analysis)
  - Otevřít v PAST (multivar → Cluster Analysis)
  - Dendrogram ukáže vztah mezi tvary

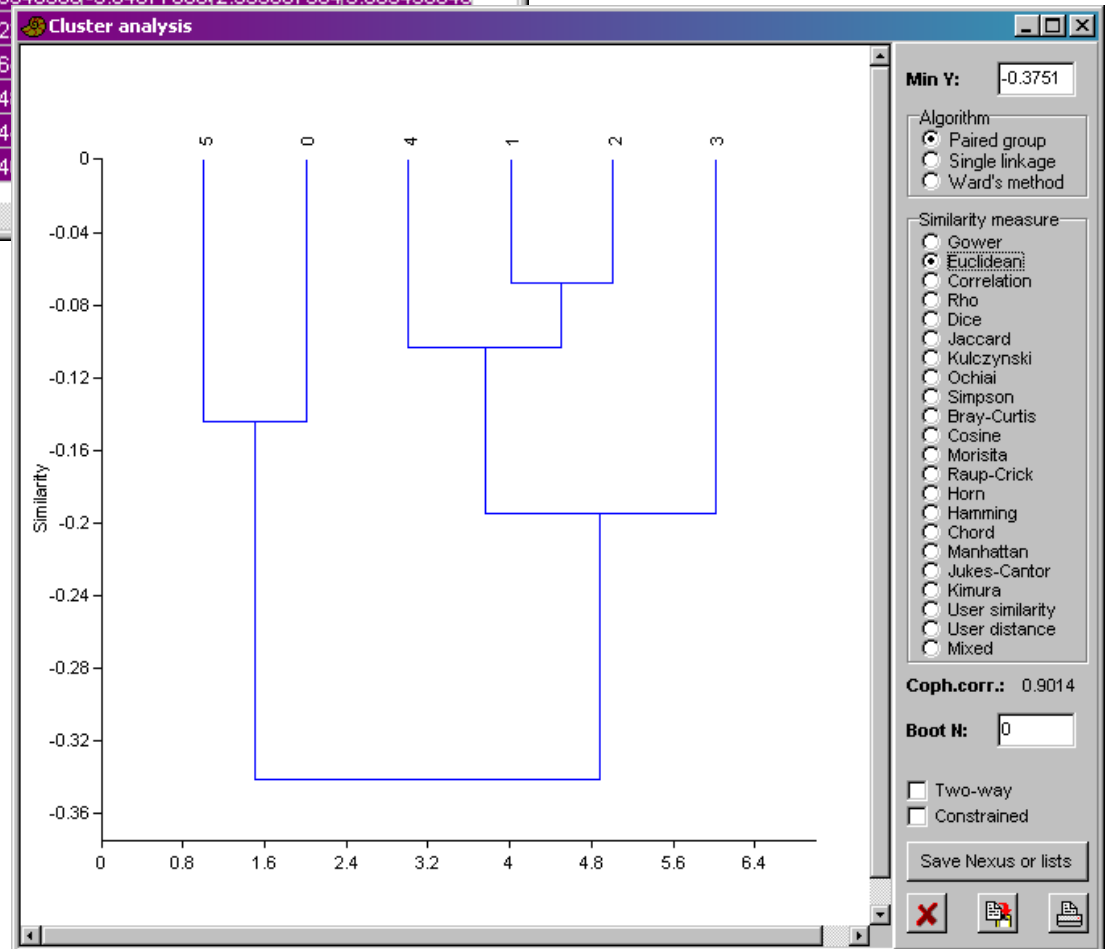
# Shluková analýza

priAl.tps

File Edit Transform Plot Statistics Multivar Model Diversity Time Geomet Strat Cladistics Script

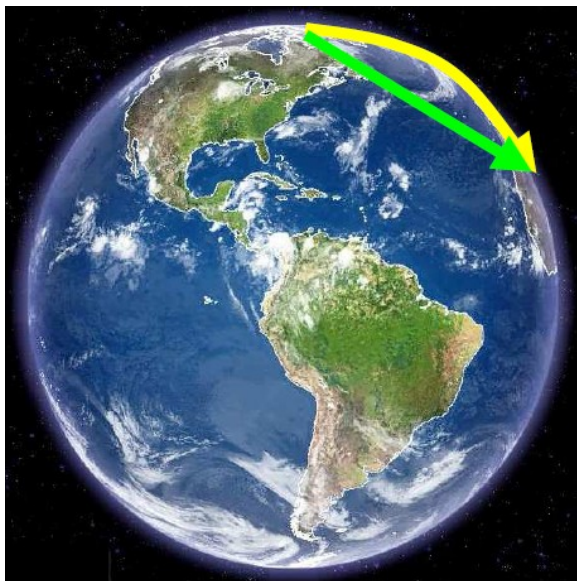
Edit mode  Edit labels  Square mode

	X1	Y1	X2	Y2	X3	Y3	X4	Y4	I
0	-2.97860380	1.876046029	-5.18275073	-1.91070487	5.625546960	-3.64877800	2.535807584	3.683436846	
1	-1.39814203	1.700911559	-6.95239647	-1.82978629	5.532				
2	-1.62052123	2.063204667	-6.81172912	-1.82707287	5.736				
3	-2.21389550	1.481798003	-6.51153690	-1.00155544	6.644				
4	-8.22724981	1.714575637	-7.40275865	-1.82544948	5.604				
5	-2.41927769	2.245326855	-5.62001940	-1.91530523	5.914				



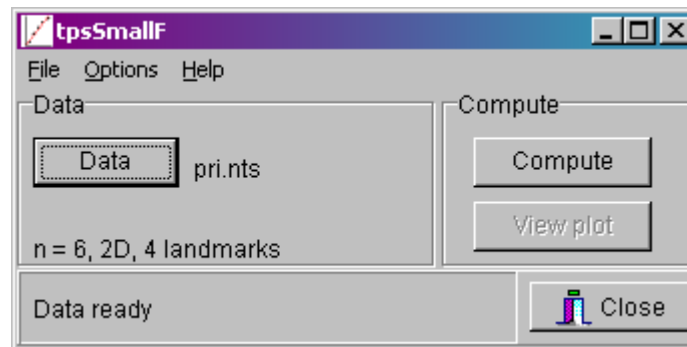
# Prostor tvarů

- Jak poznám, že množinu tvarů lze spolehlivě porovnávat a vzájemně měřit?
- Software: tpsSmall
  - Měří odchylku procrustes distance a euklidian distance
- Nutné ověřit, pokaždé když porovnávám tvar

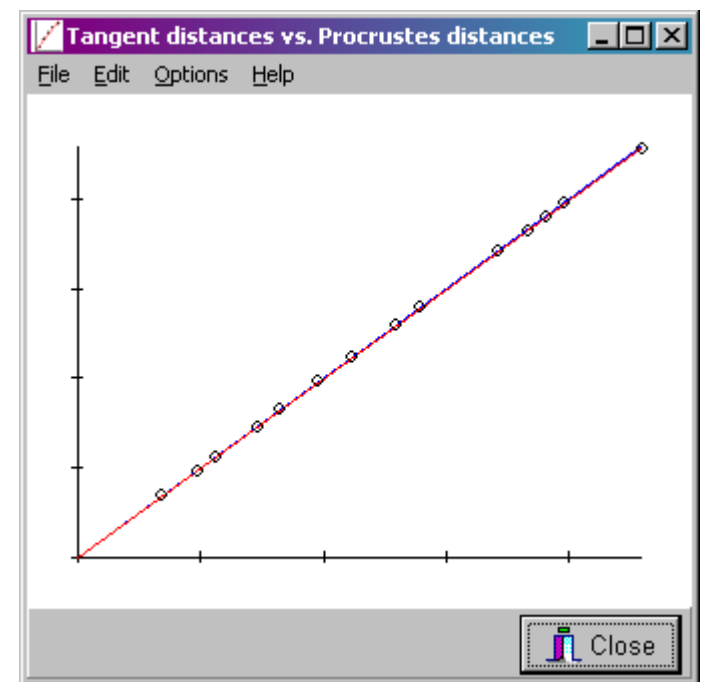


# Příklad 4

- Ověřit, že je možné porovnávat `data/2D` distances
- tpsSmall → nahrát data → výpočet → zobrazit

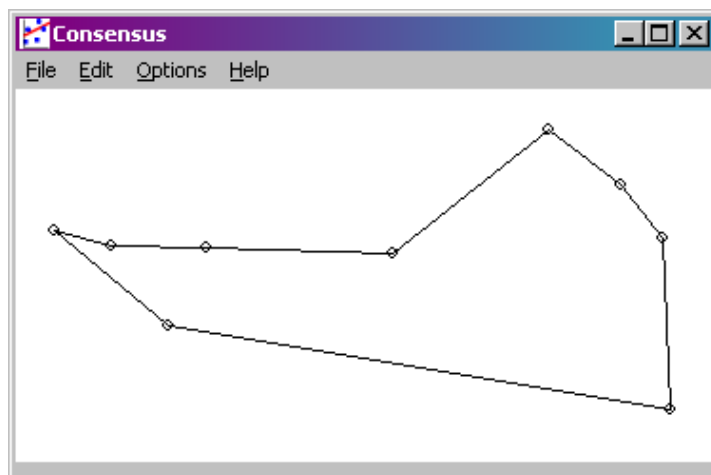


- Vyzkoušet i na `data/2D` ontogeny



# Průměrný tvar

- Consensus, mean shape
- Rozptyl
- tpsSmall/tps... File → Save → Consensus...
- tpsRegr Display → Consensus
- Excel



Seřit1 - Microsoft Excel

Domů Vložení Rozložení stránky Vzorcy Data Revize Zobrazení

Calibri 11

Vložit

Schránka Písmo Zarovnání Číslo

B7 =PRŮMĚR(B1:B6)

	A	B	C	D	E	F	G	H	I	J
1	280	369	268	158	777	307	484	569		
2	414	462	239	190	792	455	600	553		
3	387	467	279	199	751	498	568	539		
4	442	490	282	254	978	556	674	578		
5	523	469	322	230	816	454	661	508		
6	151	350	54	154	535	198	333	407		
7	366.1667	434.5	240.6667	197.5	774.8333	411.3333	553.3333	525.6667		
8										
9										
10										

Vyberte cíl a stisk... Průměr: 448.2619048 Počet: 7 Součet: 3137.833333 100 %

- Použití: shrnutí celkové podobnosti více jedinců



# Popis tvaru

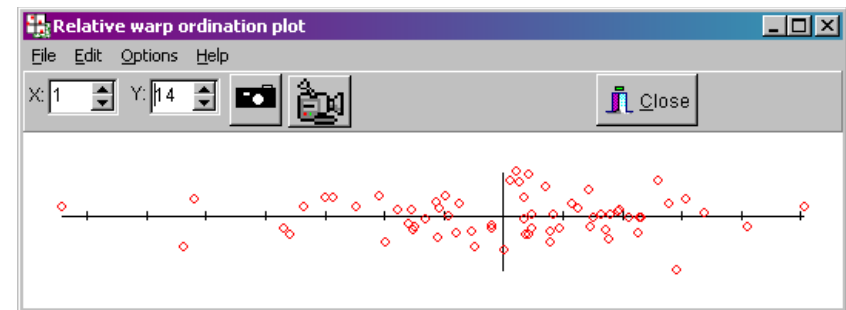
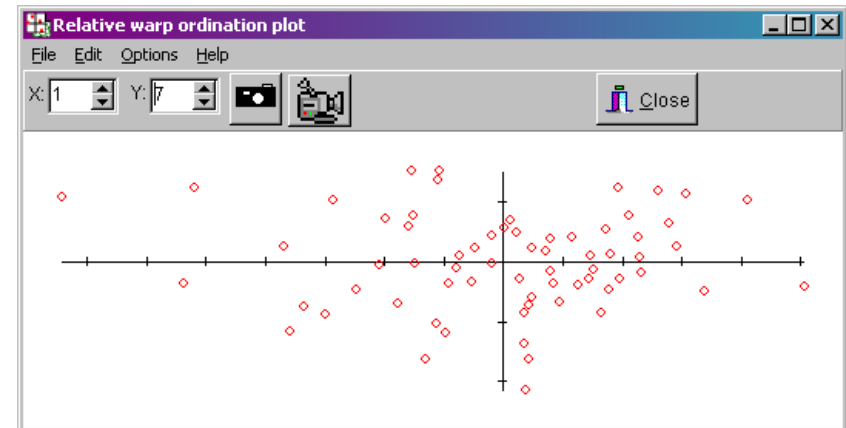
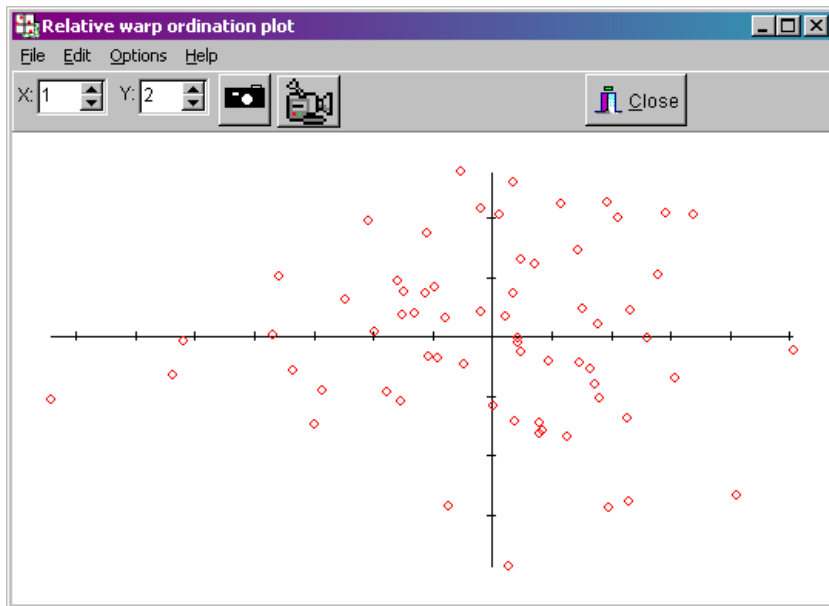
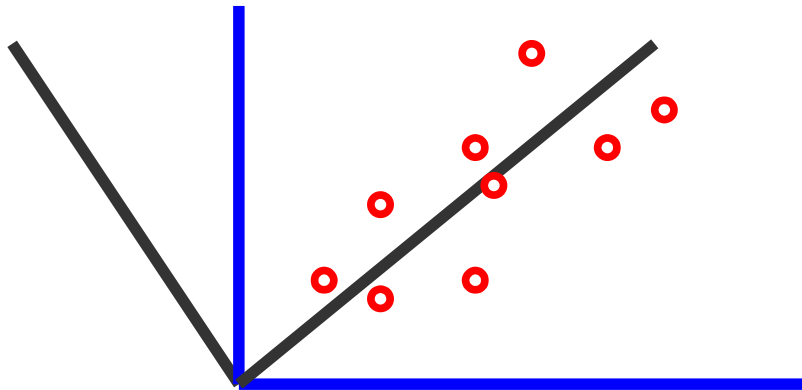
- Jak popsat složitý tvar menším počtem čísel než je počet landmarků, který tvar popisuje?
- Statisticky analyzovat množinu jedinců a najít hlavní trendy rozdílů.
- Matematický nástroj:
  - GPA – prokrustovská analýza
  - PCA – analýza hlavních komponent
- Software:
  - tpsRelw
  - PAST

# PCA



- Hledání nové báze pro data
  - $\text{data} = \{[x_1, y_1, x_2, y_2, \dots], [x_1, y_1, x_2, y_2, \dots], \dots\} = \{\mathbf{v}_1, \mathbf{v}_2, \dots\}$
  - $\mathbf{v}_1 = x_1 \cdot \mathbf{e}_1 + y_1 \cdot \mathbf{e}_2 + \dots$ ,  $\mathbf{e}_1 = [1, 0, 0, \dots]$ ,  $\mathbf{e}_2 = [0, 1, 0, \dots]$
  - Hledám nové  $\mathbf{e}_1$  tak aby  $x_1$  přes všechny  $\mathbf{v}$  měly největší rozptyl
- Matematicky: hledají se vlastní čísla/vektory kovarianční matice reziduí vstupních dat
  - Vlastní vektory jsou nové  $\mathbf{e}_1$ ,  $\mathbf{e}_2$
  - Jsou potřeba i nové  $x_1$ ,  $y_1$ , ... (souřadnice), abychom dohromady dostali  $\mathbf{v}_1$ ,  $\mathbf{v}_2$
  - Vlastní číslo odráží poměrné množství informace ve směru  $\mathbf{e}_i$

# PCA obrazem



# Příklad 5



- tpsRelw
  - Načíst data, data/2D means, data/2D distances
  - File → Open links, Consensus, Partial warp, Relative warp
  - Display → Relative warp
- PAST
  - Načíst data
  - Transform → Procrustes analysis, Multivar → Principal components
- Jde o trochu odlišné věci

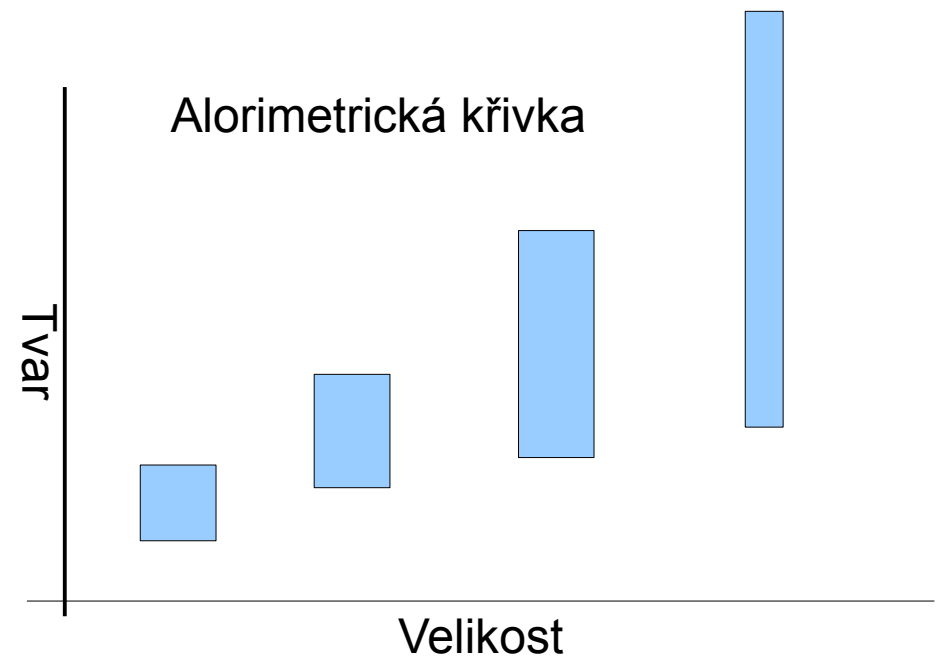
# Warps



- Principal warps
  - Stará terminologie
- Partial warps
  - GPA souboru dat  $\rightarrow$  mean shape
  - TPS mean shape na konkrétního jedince  $\rightarrow$  matice  $W$   
parametrů TPS  $\rightarrow$  bending energy matrix  $L^{-1}$
  - vlastní vektory matice  $L^{-1}$  jsou **partial warps**
- Relative warps
  - PCA z partial warp scores (souřadnic jedince vzhledem k partial warps bázi)
  - Partial warps jsou přenásobené faktorem  $\alpha$

# Vztah velikosti a tvaru

- Regrese mezi velikostí (nezavislá) a tvarem (závislá)
- Alometrie, Vývoj (věk  $\rightarrow$  tvar), Růst (věk  $\rightarrow$  velikost)
- Závislost mnoha parametrů (tvar je definován mnoha parametry) na jednom (centroidová velikost)
- Matematický nástroj: Multivarietní regrese
- Software: tpsRegr, PAST

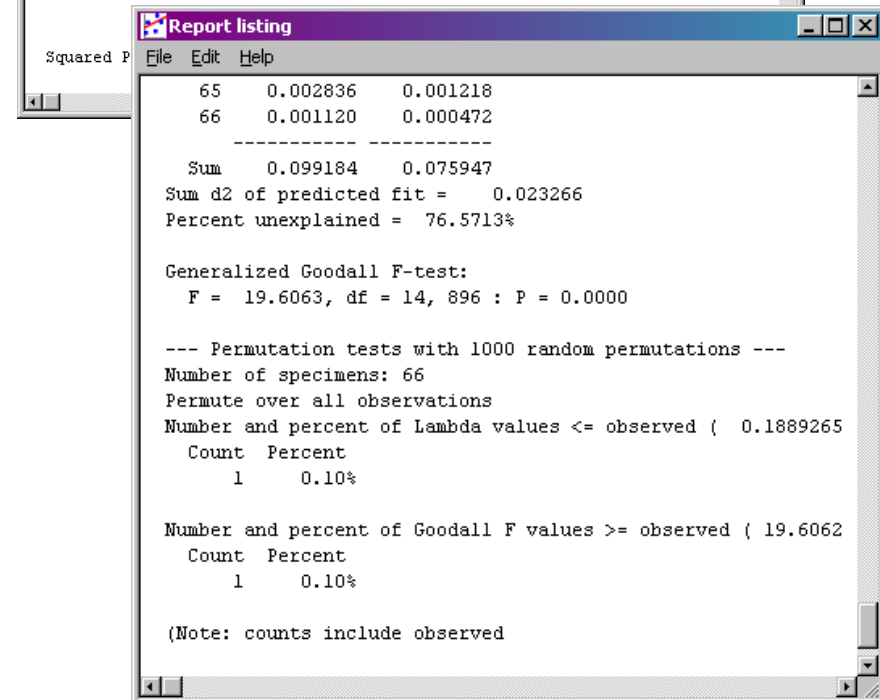
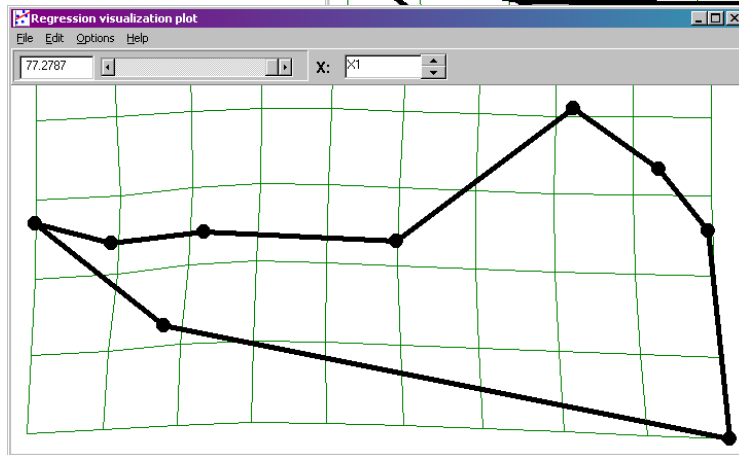
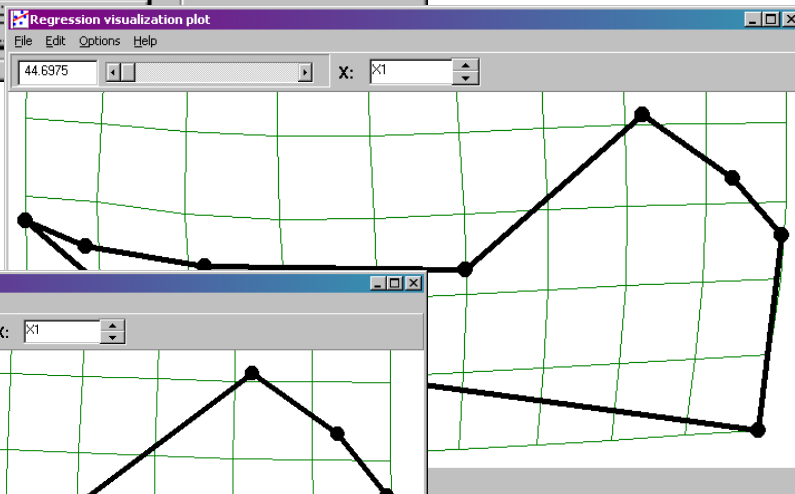
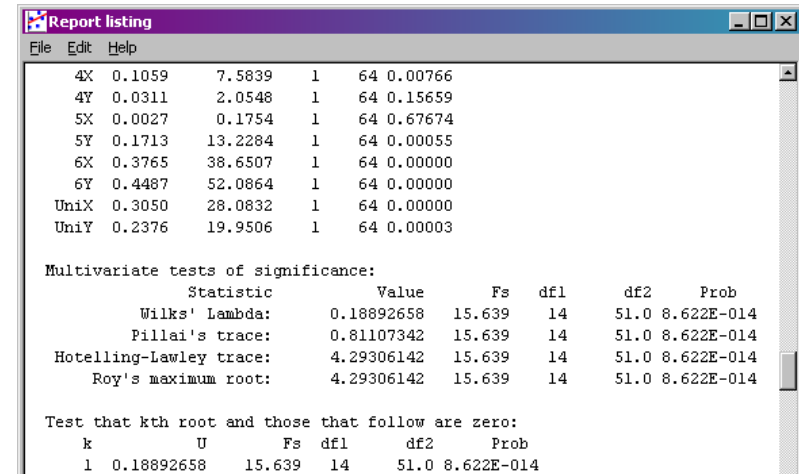
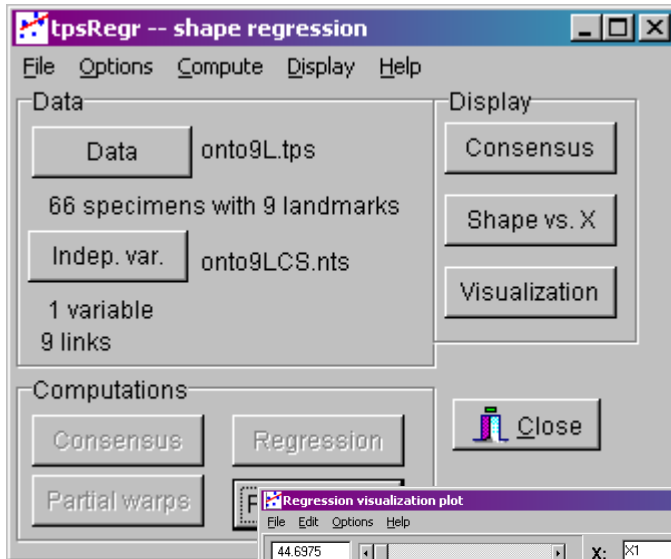


# Příklad 6



- **Data** `data/2D onthogeny`
  - `links.nts`, pro vizualizaci
  - Centroidová velikost, například v `tpsSmall`, `tpsRelw`
- Spustit `tpsRegr`
  - Otevřít data s landmarky (tvar), závislé proměnné
  - Centroidová velikost, nezávislá proměnná
  - Nahrát `links`, pro vizualizaci
  - Consensus, Partial warps, Regression, Perm.tests
  - File → View Report (číselné výsledky regrese)
    - P hodnoty odrážejí pravděpodobnost, že není vztah mezi závislými a nezávislými proměnnými
- Vizualizace

# Ukázka





# Příklad 6 pokračování

- Kontrola: Příklad lze celý provést v PAST
  - Nahrát data
  - CS, GPA, PCA
  - Model  $\rightarrow$  Linear 1 indep, n dep
- Vizualizace složitější
  - Vypočítat hodnotu regrese (v PCA souřadnicích)
  - Vypočítat skutečné souřadnice
  - Zobrazit

# Závěr

---

- Některé funkce programu PAST
- tpsRelw, tpsRegr, tpsUtil, tpsDig(2), tpsSmall
- Excel
- 4 příklady

# Q & A

---



# Reference

---



- [Andrea Cardini: GMM\\_2008\\_York](#)
- [Zelditch, Swiderski, Sheets, Fink: Geometric Morphometrics for Biologist: A Primer](#)
- [Wikipedia: The Free Encyclopedia](#)