Bi8352: Metody antropologie II
jaro 2020
Mgr. Mikoláš Jurda, Ph.D.

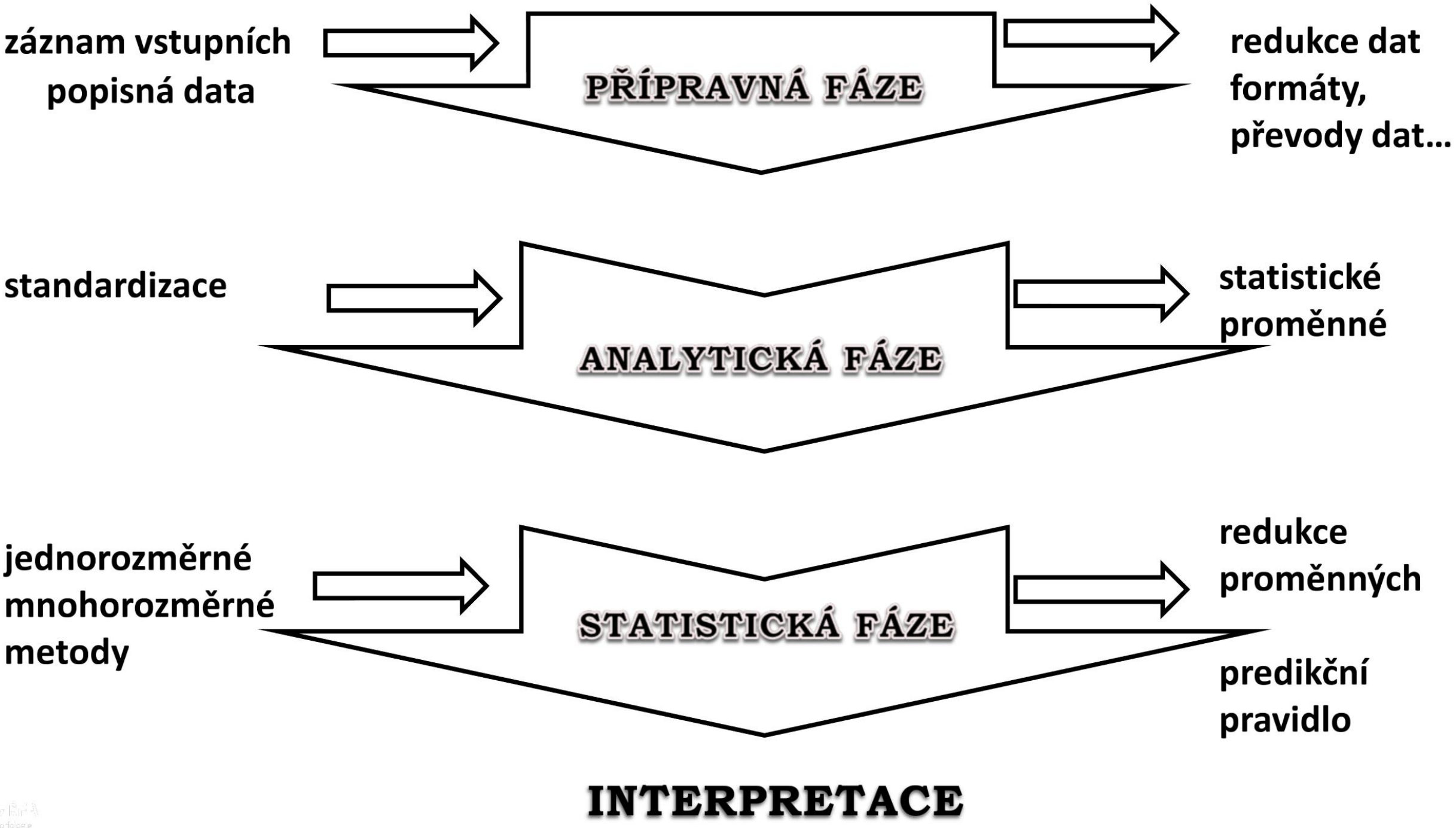
MUNI
SCI

Základy Statisticy

Proč aplikujeme statistické postupy???

Poskytují vědecký základ

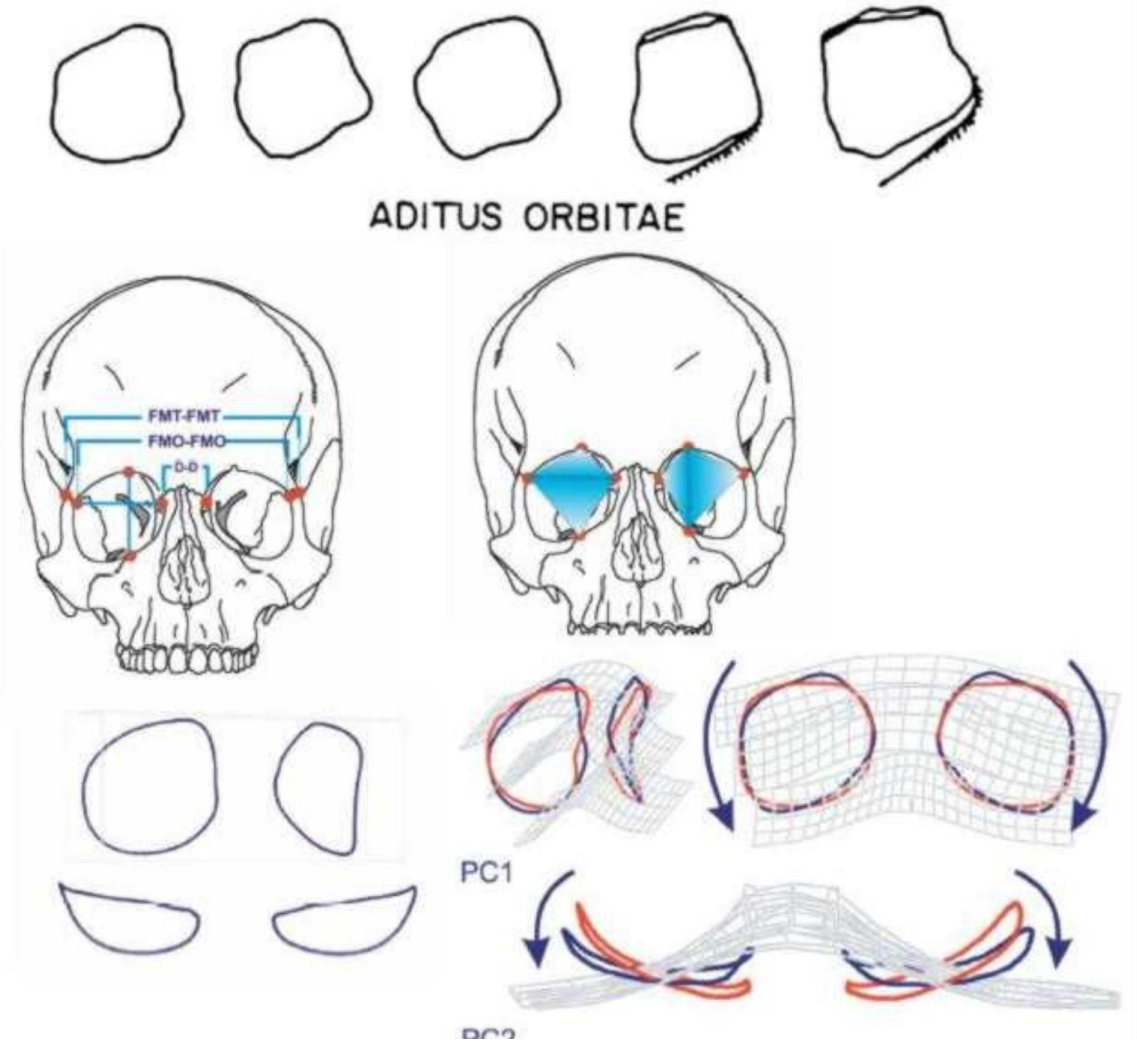
- **objektivní** sumarizace výsledků!!!
- vytváření **predikčního pravidla** – odhad neznámých vlastností na základě známých vlastností
- kombinace odlišných biologických vlastností do jednotného metodického postupu
- zjišťování chyby odhadu/určení


záznam vstupních
popisná data
PŘÍPRAVNÁ FÁZE
redukce dat
formáty,
převody dat…
standardizace
ANALYTICKÁ FÁZE
statistické
proměnné
jednorozměrné
mnohorozměrné
metody
STATISTICKÁ FÁZE
redukce
proměnných
predikční
pravidlo
INTERPRETACE

Morfoskopický přístup

Tradiční morfometrický přístup

Pokročilý morfometrický přístup



rozdílné v **rozšíření**, **využití**, **zpracování** a **možnostech**

Popisná (deskriptivní) statistika

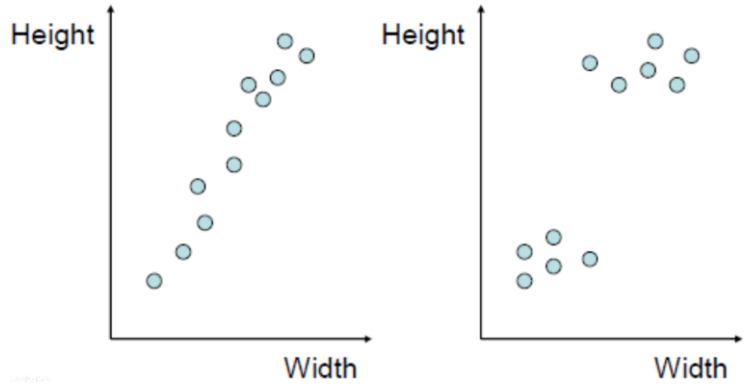
- **základní informace** o vlastnostech studovaného souboru a vztazích různých souborů a dat

- kontrola splnění předpokladů statistických testů

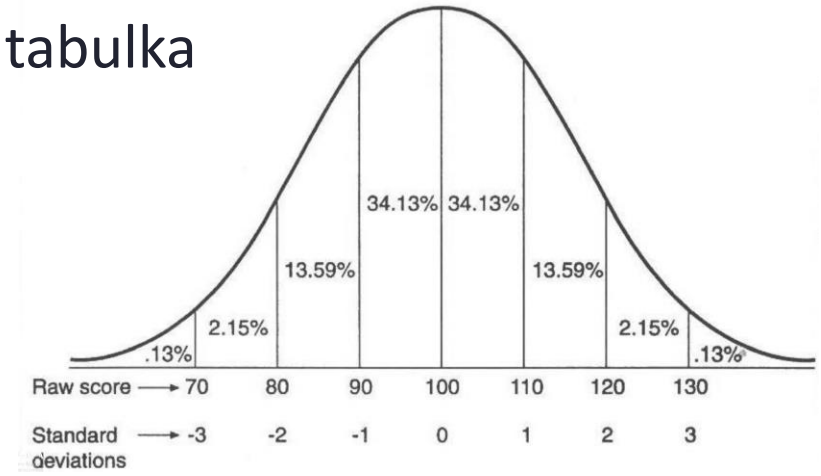
průměr
rozptyl
medián
modus
SD
směrodatná chyba
koeficient variance
normalita rozložení

histogram
frekvenční tabulka

kontingenční tabulka



korelace proměnných

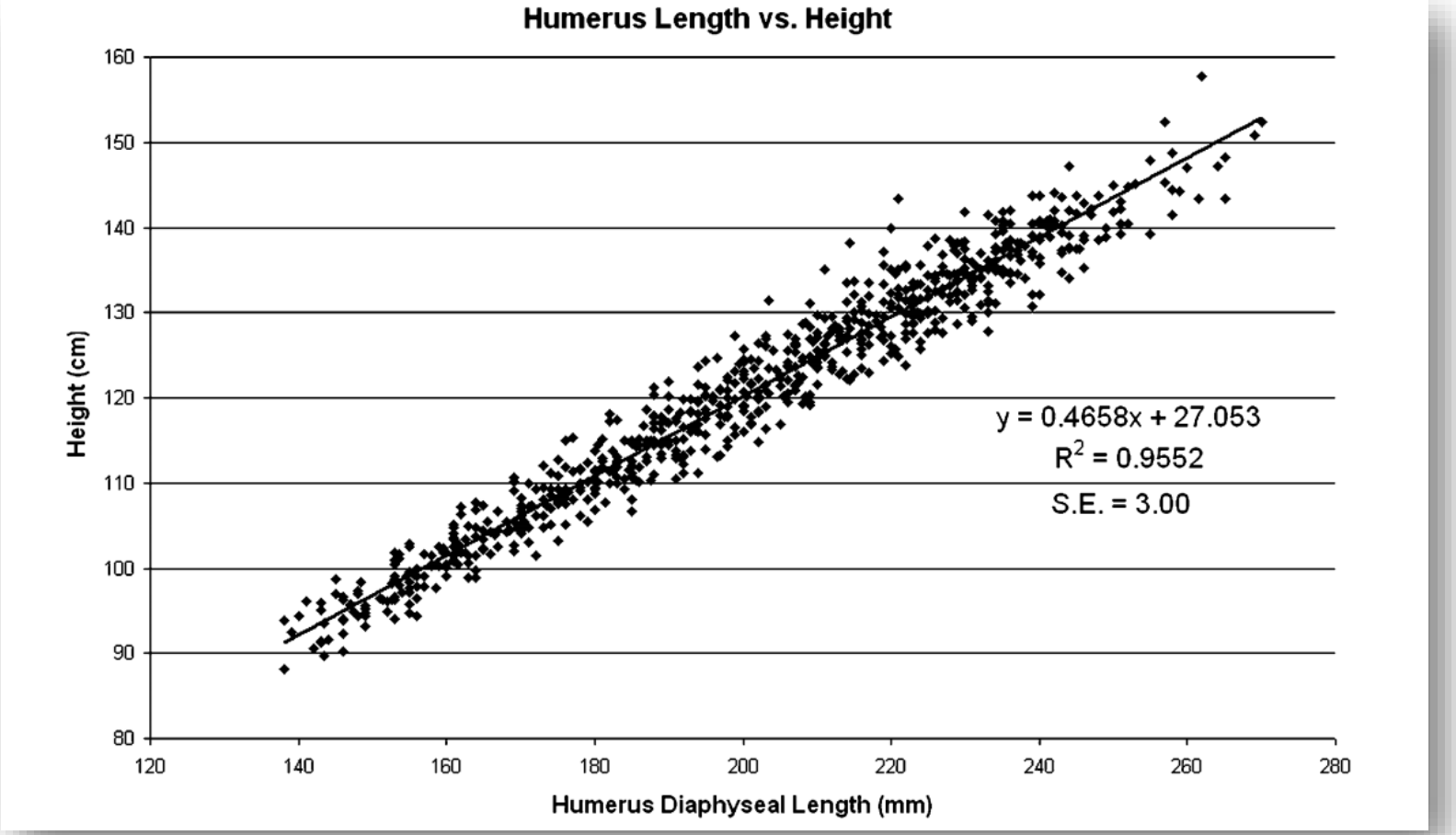


jednorozměrné metody	vícerozměrné metody	
lineární regrese	**diskriminační analýza**	**kanonická analýza**
jednorozměrná vícenásobná vícerozměrná vícenásobná vícerozměrná		
ODHAD TĚLESNÝCH PROPORCÍ	KLASIFIKACE DO JEDNÉ ZE DVOU SKUPIN	KLASIFIKACE DO JEDNÉ Z VÍCE SKUPIN
ODHAD VĚKU JEDINCE		
DALŠÍ KVANTITATIVNÍ ODHADY	KLASIFIKACE DO JEDNÉ Z VÍCE SKUPIN	JAKÉKOLIV ZAŘAZENÍ DO TYPŮ

Jednorozměrná lineární regrese

jedna vstupní proměnná
nezávislá = predikující = známá

jedna výstupní proměnná (numerická vlastnost)
závislá = predikovaná = neznámá



(Smith 2007)

Jednorozměrná lineární regrese

REGRESE **není** **KORELACE**

vyjadřuje, jak lze z nezávislé proměnné odhadnout závislou proměnnou

regrese vyjadřuje vliv změny hodnoty známé proměnné na hodnotu neznámé proměnné

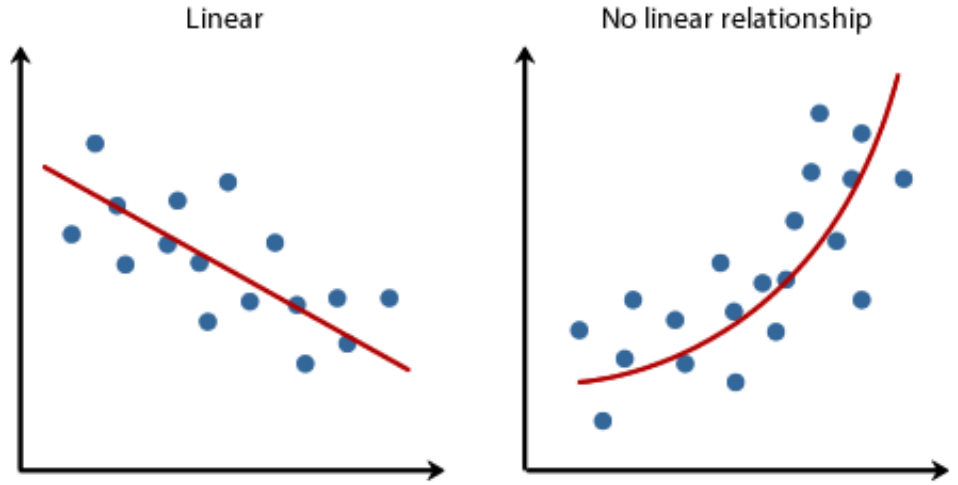
vyjadřuje vztah mezi dvěma rovnocennými proměnnými

vypovídá o tom, do jaké míry se dvě proměnné mění společně

Jednorozměrná lineární regrese

jedna vstupní proměnná
nezávislá = predikující = známá

jedna výstupní proměnná (numerická vlastnost)
závislá = predikovaná = neznámá

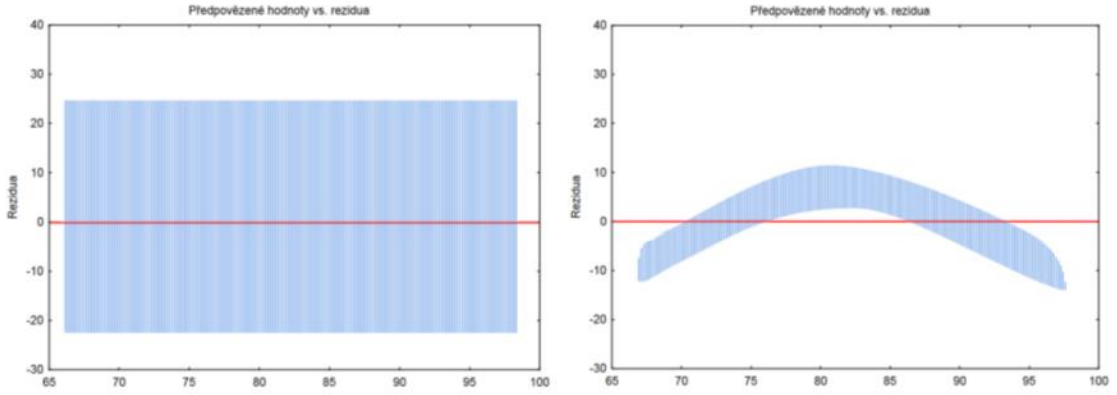


Předpoklady

- vztah je lineární
- data získána nezávisle na sobě

Předpoklady výstupů

- střední hodnota chybové složky je 0
- chybová složka má konstantní rozptyl
- jednotlivé složky chybového vektoru jsou nekorelované
- reziduální složka má normální rozdělení



Jednorozměrná lineární regrese

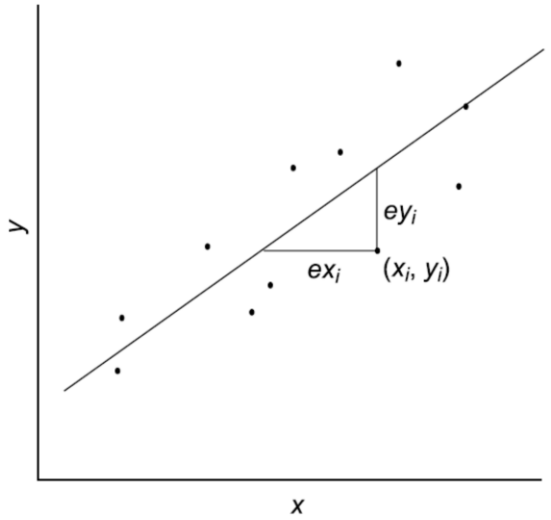
jedna vstupní proměnná
nezávislá = predikující = známá

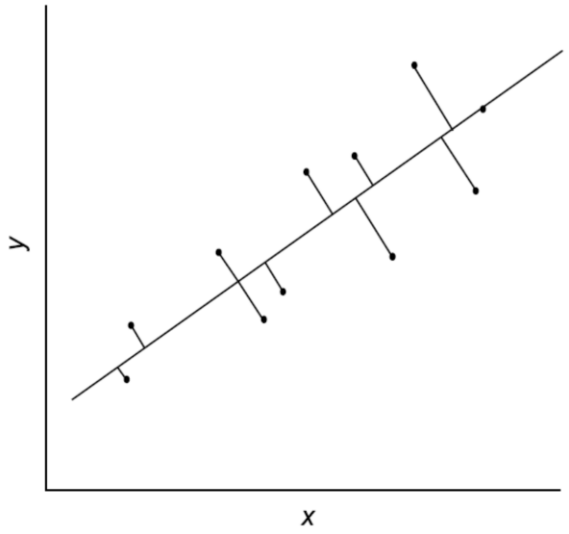
jedna výstupní proměnná (numerická vlastnost)
závislá = predikovaná = neznámá

y = ax b + (+E)

různé metody vyjádření chyby
(jak proložit body přímku?)

- metoda nejmenších čtverců
  - pouze y-proměnné
  - x i y (RMA)
- euklidovská vzdálenost





(RMA – reduced major axis)

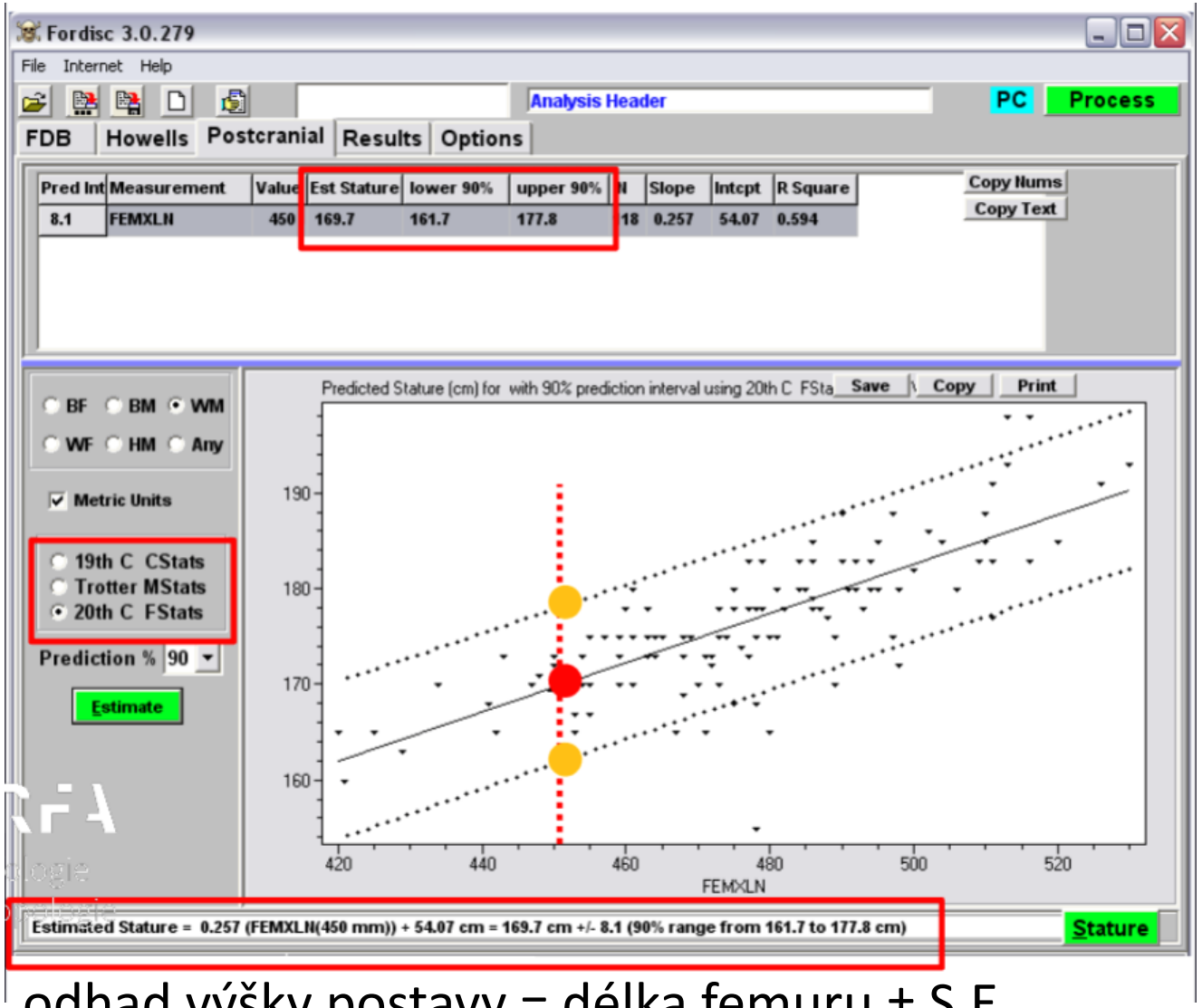
Jednorozměrná lineární regrese

Výstupy $y = ax + b (+E)$

predikční pravidlo = lineární rovnice
- hodnota závislé proměnné (y)
- koeficient závislosti (a)
- položení v prostoru (b)

interval spolehlivosti = konfidenční interval

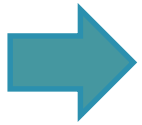
standardní chyba odhadu



odhad výšky postavy = délka femuru ± S.E.

Vícenásobná lineární regrese

více vstupních proměnných
nezávislá = predikující = známá



jedna výstupní proměnná (numerická vlastnost)
závislá = predikovaná = neznámá

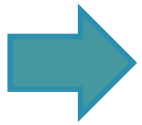
y = ax + bx + c (+E)

+

- vstupní proměnné by neměly korelovat

Vícerozměrná lineární regrese

jedna vstupní proměnná
nezávislá = predikující = známá



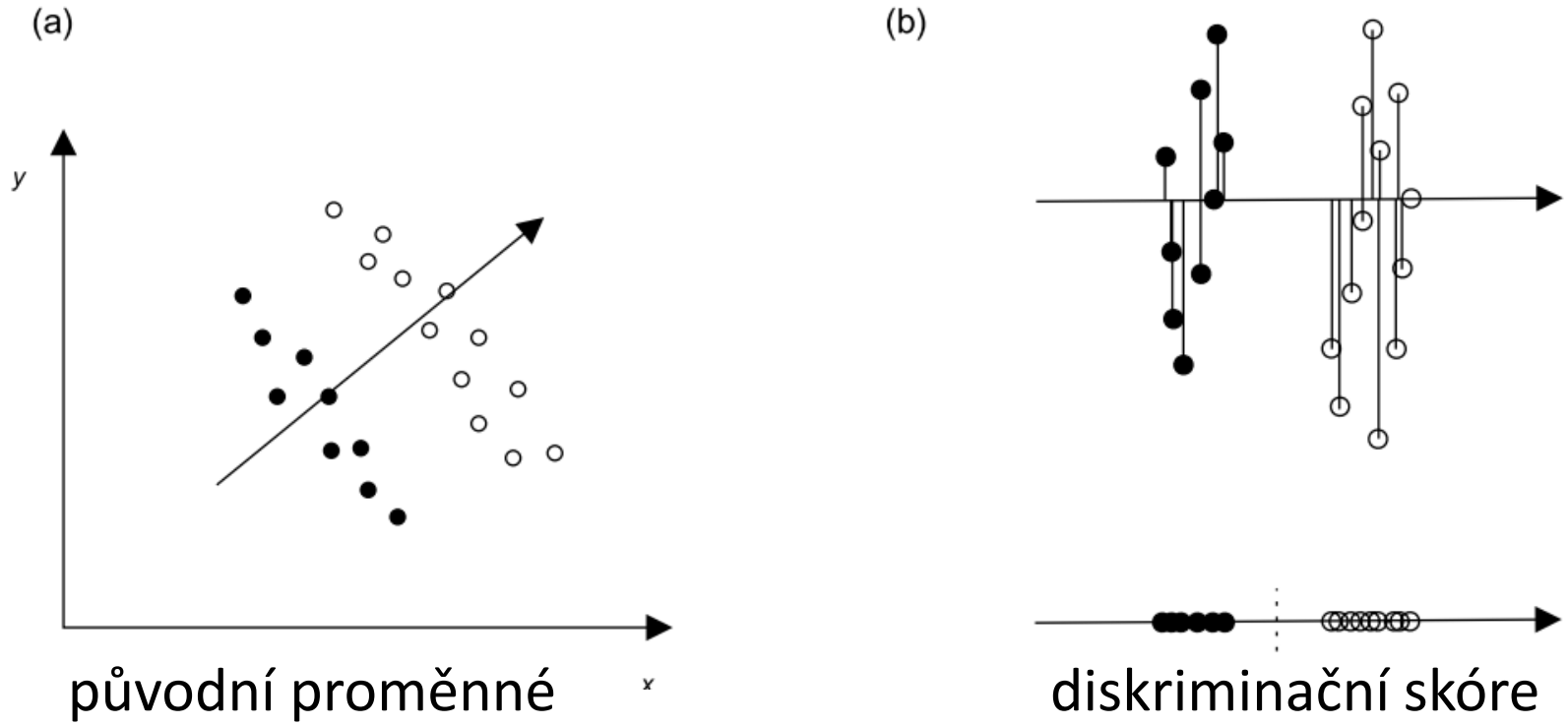
dvě a více výstupních proměnných (numerická vlastnost)
závislá = predikovaná = neznámá

Diskriminační analýza

minimálně dvě nezávislé proměnné

predikční model pro odlišení mezi dvěma skupinami

Variabilita proměnných je zpracována s ohledem na předem dané (a priori známé) rozdělení do skupin.



původní proměnné

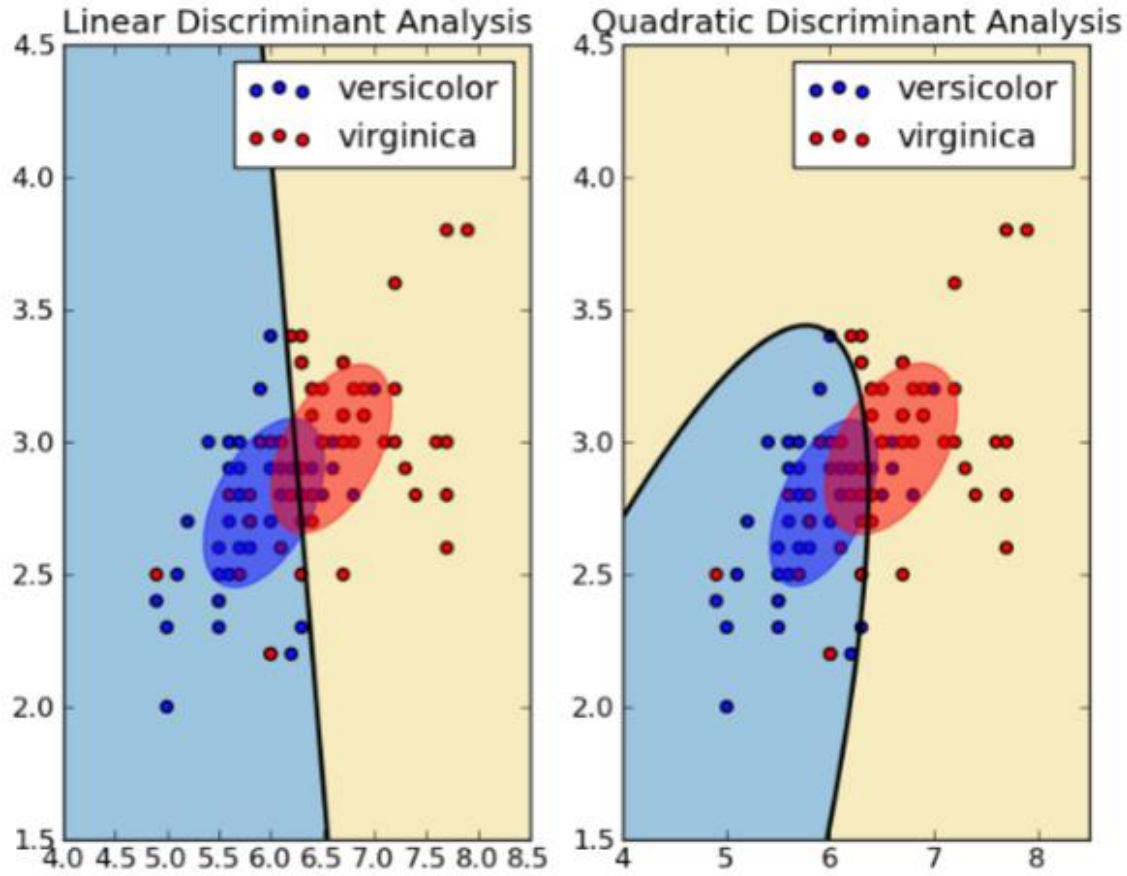
diskriminační skóre

Diskriminační analýza

minimálně dvě nezávislé proměnné

predikční model pro odlišení mezi dvěma skupinami

- založeno na **lineárním modelu.**
- diskriminační skóre – lineární kombinace původních proměnných



Diskriminační analýza

Výstupy:

- **predikční pravidlo**
- diskriminační skóre pro každý případ
- **nestandardizované koeficienty pro každou proměnnou (použity v diskriminační rovnici**)
- standardizované koeficienty (vyjadřují podíl dané veličiny na diskriminačním skóre)
- spolehlivost pravidla
- Mahalanobisova vzdálenost
- aposteriorní pravděpodobnost
- spolehlivost klasifikace

White-Indian:
3.05(*Basion-prosthion*) — 1.04(*Glabello-occipital length*) — 5.41(*Maximum width*) + 4.29(*Basion-bregma height*) — 4.02(*Basion-nasion*) + 5.62(*Maximum diameter bi-zygomatic*) — 1.00(*Prosthion-nasion height*) — 2.19 (*Nasal breadth*).

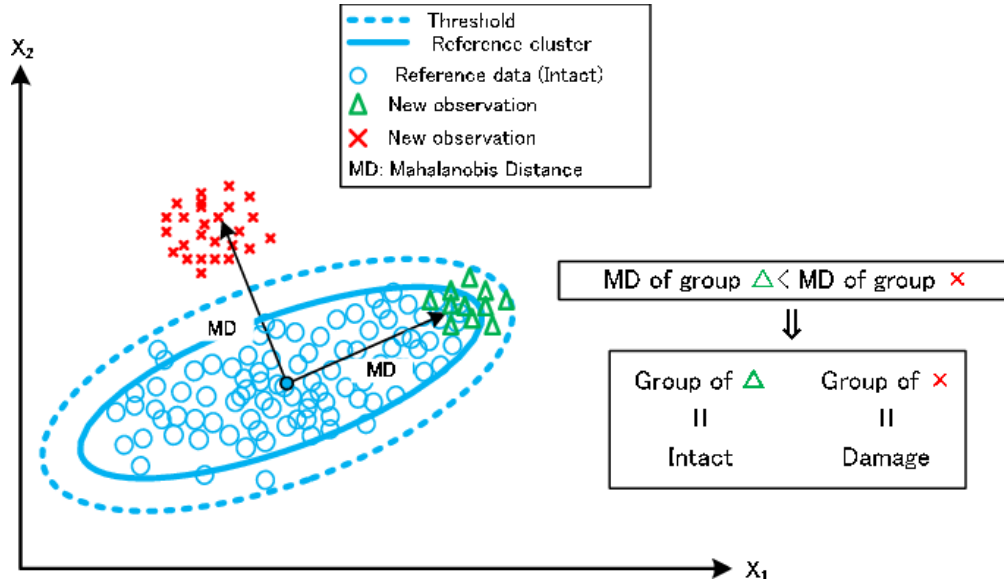
umožňuje přiřadit neznámému objektu regresní skóre a na základě jeho hodnoty jej zařadit do skupiny

Diskriminační analýza

Výstupy:

- predikční pravidlo
- diskriminační skóre pro každý případ
- nestandardizované koeficienty pro každou proměnnou (použity v diskriminační rovnici)
- standardizované koeficienty (vyjadřují podíl dané veličiny na diskriminačním skóre)
- spolehlivost pravidla
- **Mahalanobisova vzdálenost**
- **aposteriorní pravděpodobnost**
- spolehlivost klasifikace

Mahalanobisova vzdálenost
popisuje vzdálenost centroidů skupin (bere v úvahu korelaci mezi parametry a je nezávislá na jejich rozsahu)



Posterior probability – pravděpodobnost zařazení objektu do skupiny (p toho, že objekt patří do té které skupiny) – vychází z Makalanobisových vzdáleností ke skupinám a *a priori* pravděpodobnosti

Diskriminační analýza

Výstupy:

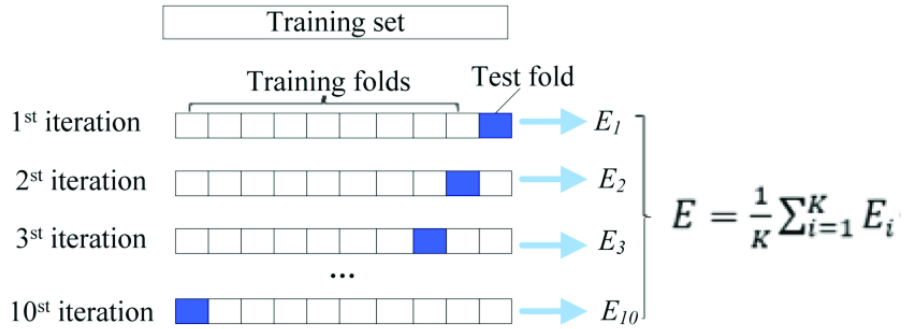
- predikční pravidlo
- diskriminační skóre pro každý případ
- nestandardizované koeficienty pro každou proměnnou (použity v diskriminační rovnici)
- standardizované koeficienty (vyjadřují podíl dané veličiny na diskriminačním skóre)
- spolehlivost pravidla
- Mahalanobisova vzdálenost
- aposteriorní pravděpodobnost
- **spolehlivost klasifikace**

Spolehlivost zařazení případů do skupin na základě predikčního pravidla

			sex_real		Total female
			Female	Male	
Team B	Estimated sex	Female	20	1	21
		Male	1	28	29
	Total		21	29	50
Team A	Estimated sex	Female	19	0	19
		Male	2	29	31
	Total		21	29	50

Both teams achieved 96% accuracy (48 of 50) in determining the correct sex classification.

resubstituce
křížová validace (cross-validace)



ještě lépe
testování na nezávislém vzorku

Kanonická analýza

minimálně tři proměnné → predikční model pro odlišení mezi více **než dvěma skupinami**

Variabilita proměnných je zpracována s ohledem na předem dané (**a priori známé**) rozdělení do skupin – nové proměnné (kanonické osy), maximalizují rozdíly mezi skupinami.

Vlastnosti popsané původními proměnnými jsou převedeny na kanonické proměnné (k-1, kde k je počet skupin)

Pro každý prvek existuje hodnota kanonické proměnné – místo, kam dopadne na kanonické ose

Výstupy:
kanonické rovnice (k-1)
CS1 = a1x1 + b1x2 + c1x3 ... + C1
CS2 = a2x1 + b2x2 + c2x3 ... + C2

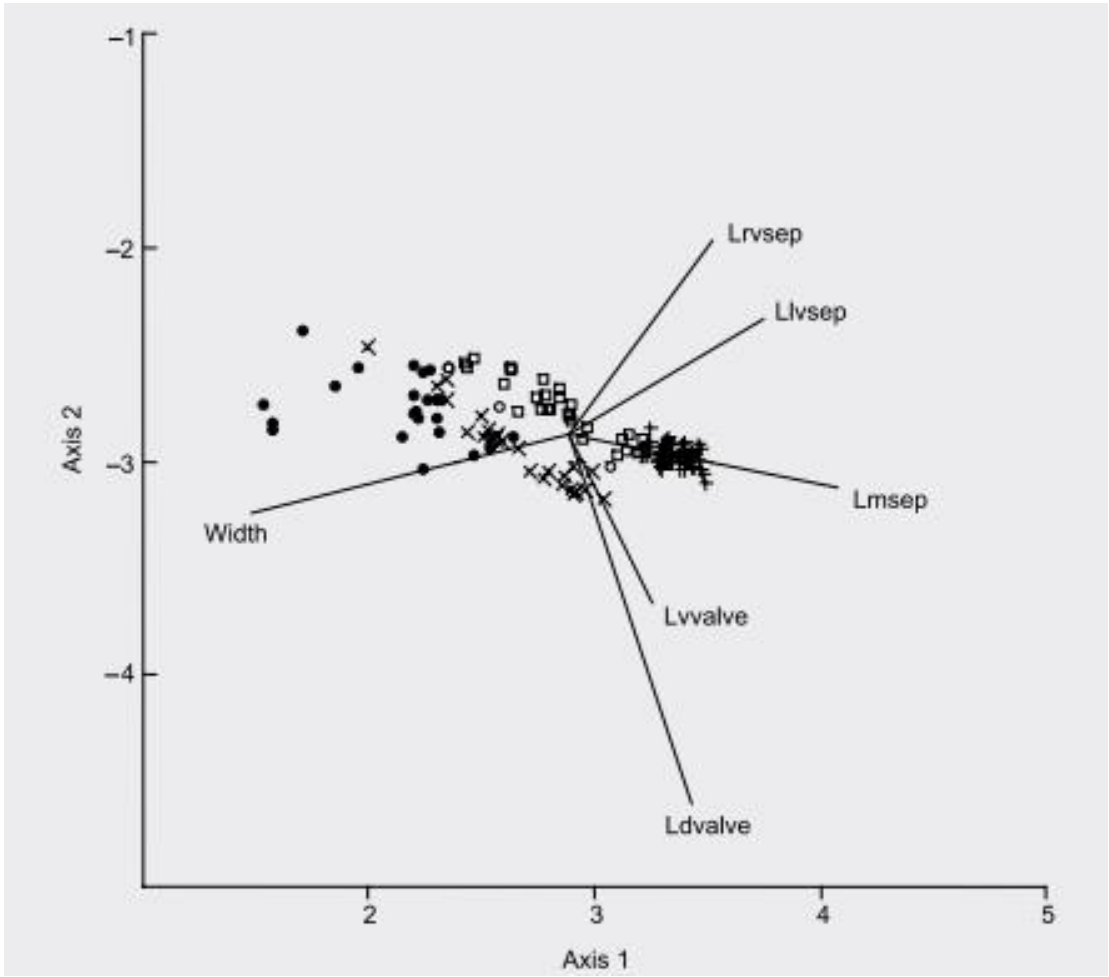
standardizované a nestandardizované koeficienty

Kanonická analýza

minimálně tři proměnné

predikční model pro odlišení mezi více **než dvěma skupinami**

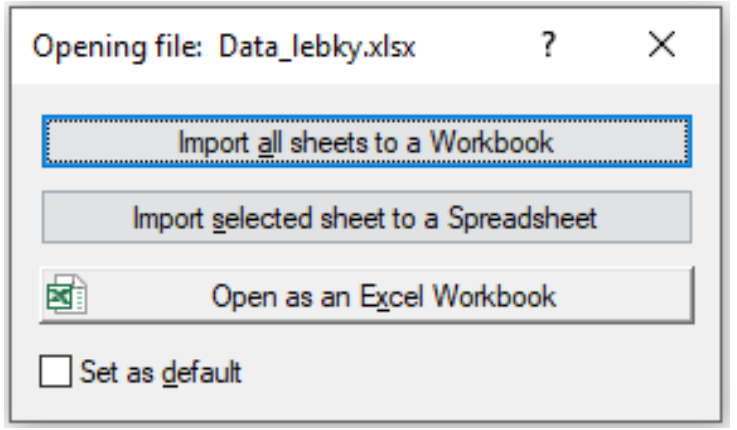
Grafy – redukce proměnných na to „podstatné“



Statistica – import souborů a práce s nimi

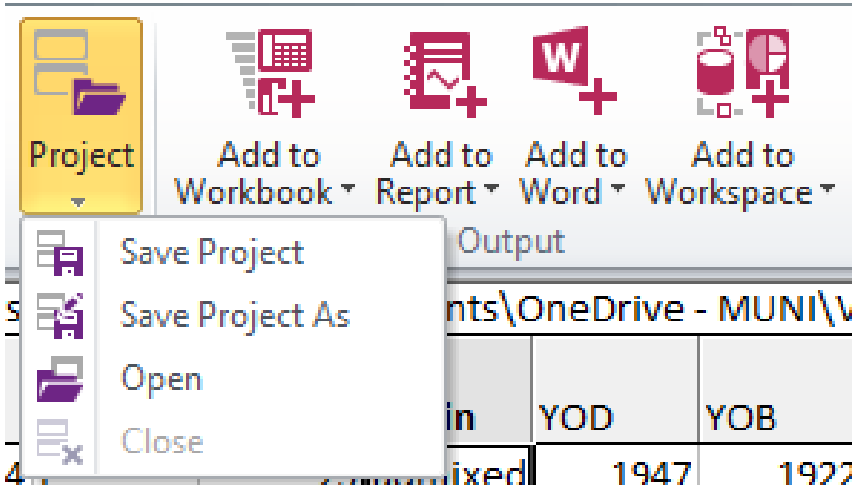
- importovat ze souborů různých typů – excel, csv, txt a také vložením ze schránky

  Práci ušetří, pokud jsou buňky v původním excelovském souboru ve správném formátu. Pokud po načtení formát proměnných neodpovídá našim požadavkům, jde formát upravit nastavením jednotlivých proměnných (dvojité poklikání na buňku s názvem proměnné) – typicky nastavení číselných proměnných na *Type* > ***Double***

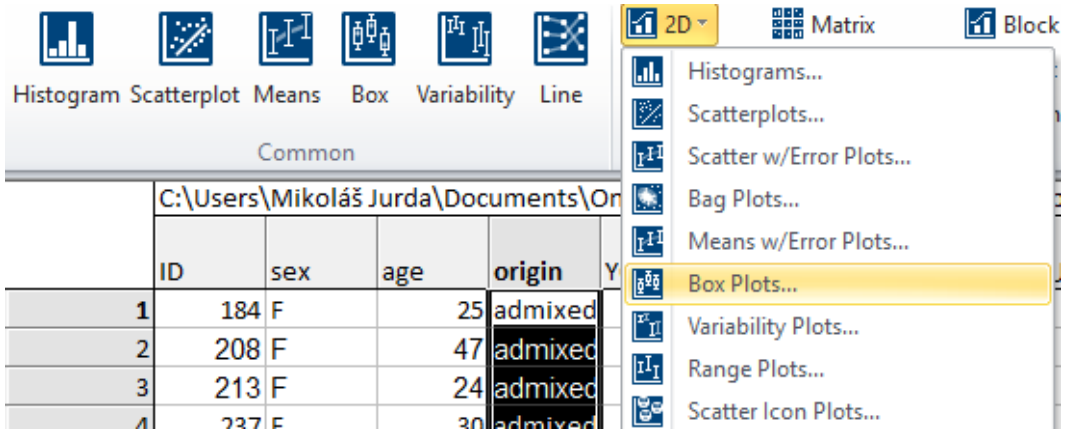


- datový soubor a výstupy analýz je možné uchovávat v různých formátech

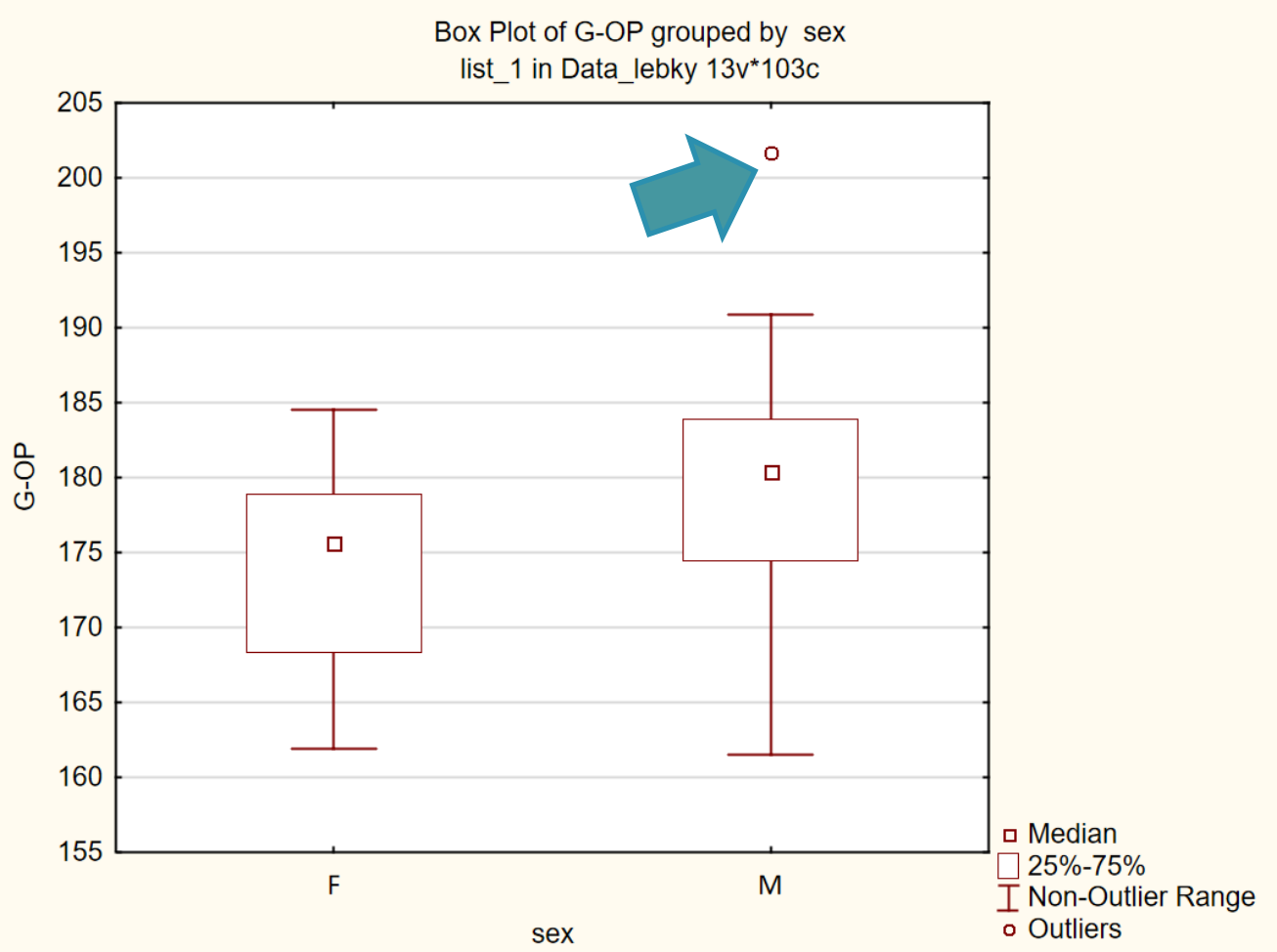
  Možnost uložit vše v přehledném stromu nabízí **Project**. Výstupy je možné exportovat také například ve formátu .doc.



Popisná statistika – vizuální hodnocení – krabicový graf

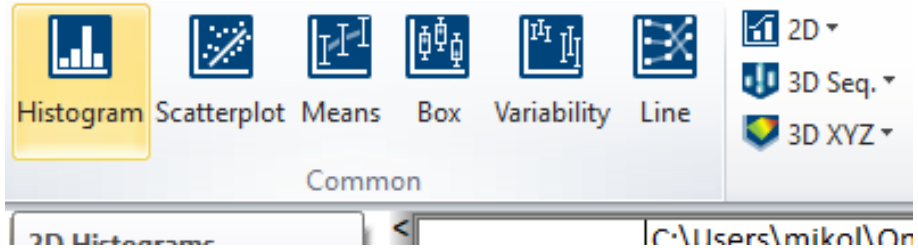


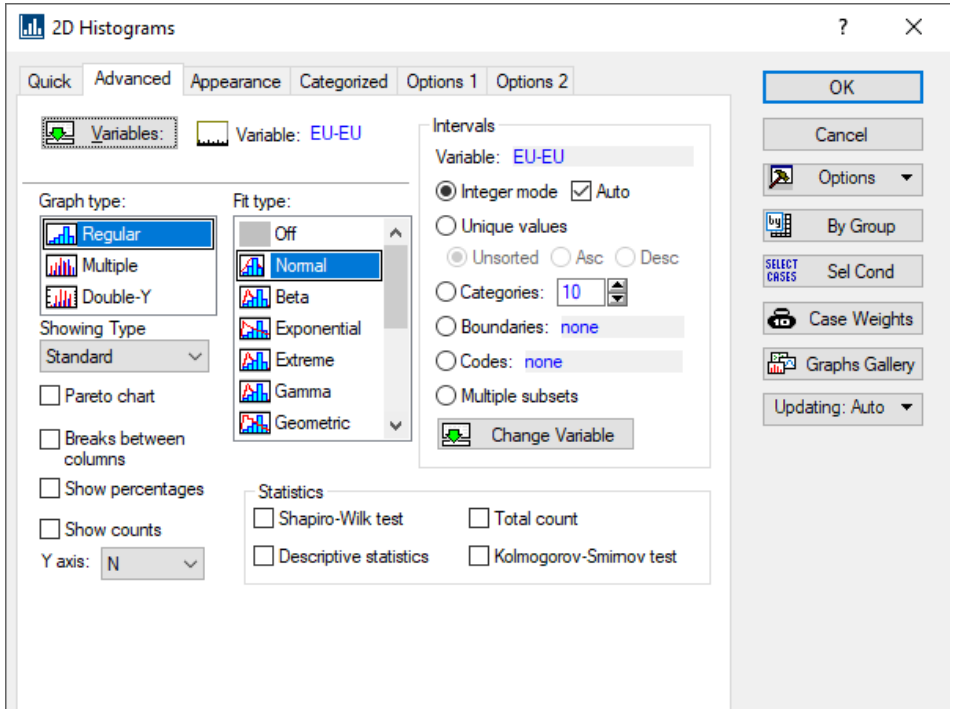
Pokud nezadáte grupovací proměnnou, zobrazí se graf pro celý soubor, pokud ano, pak odděleně pro definované

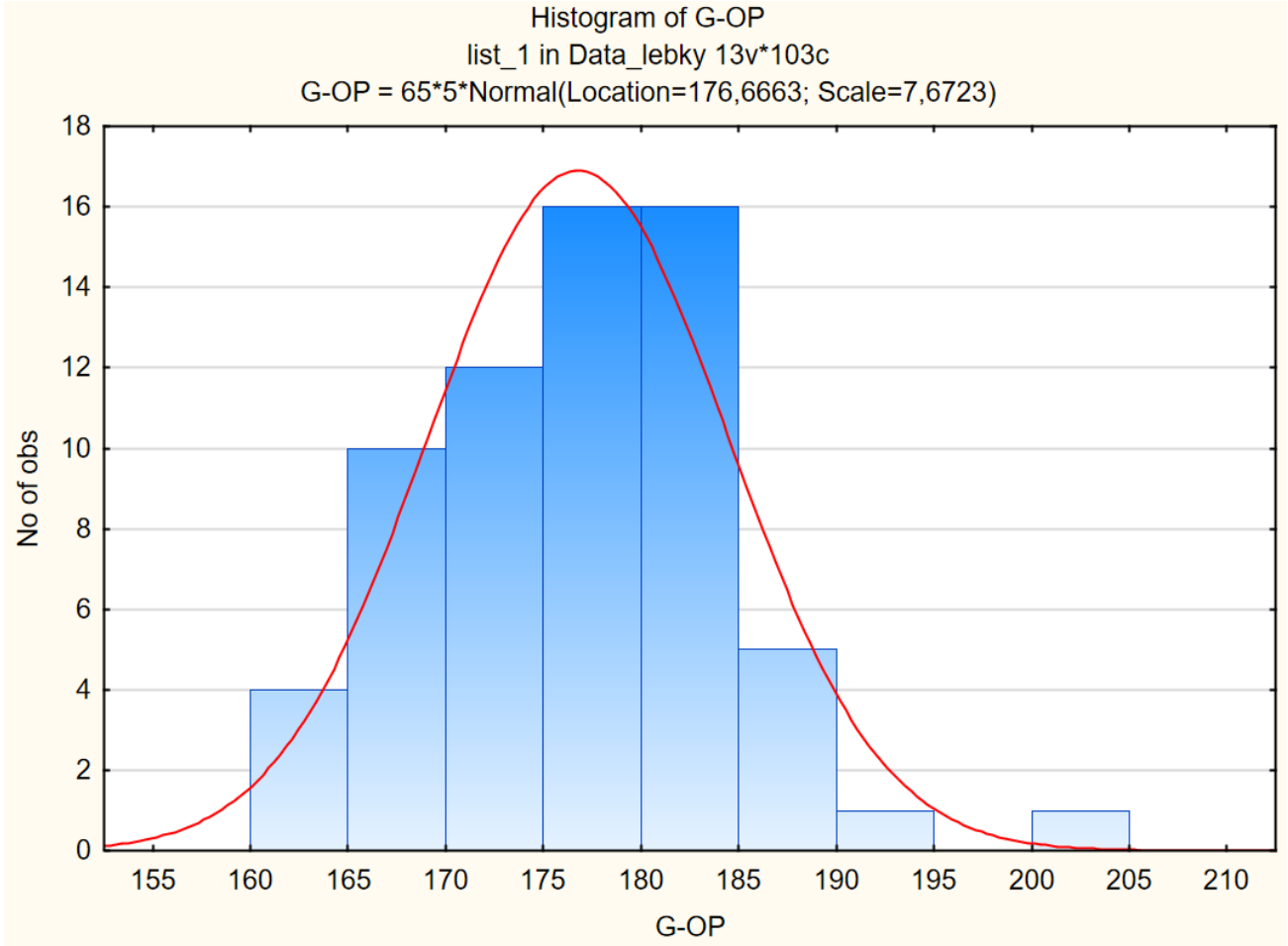


Krabicový graf pro dvě skupiny – m a f – dobrý pro vyhledávání **extrémních případů (například chyb v datech)** – při podržení myši nad odlehlou hodnotou se zobrazí její ID

Popisná statistika – vizuální hodnocení – histogram

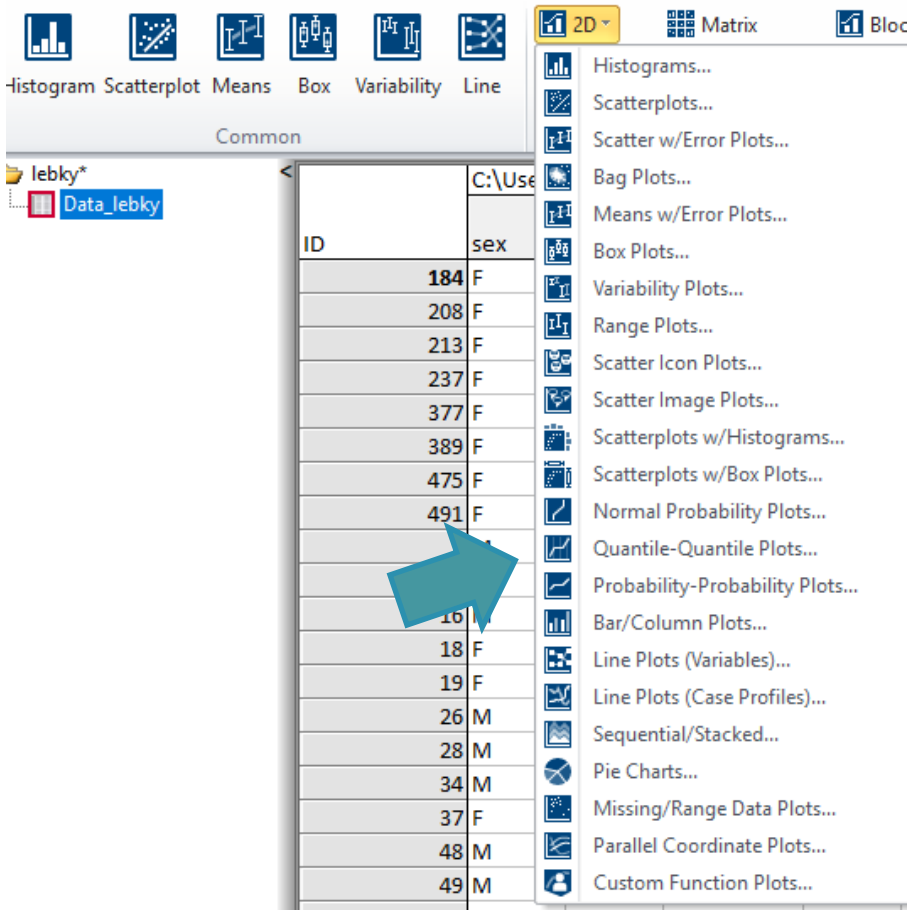


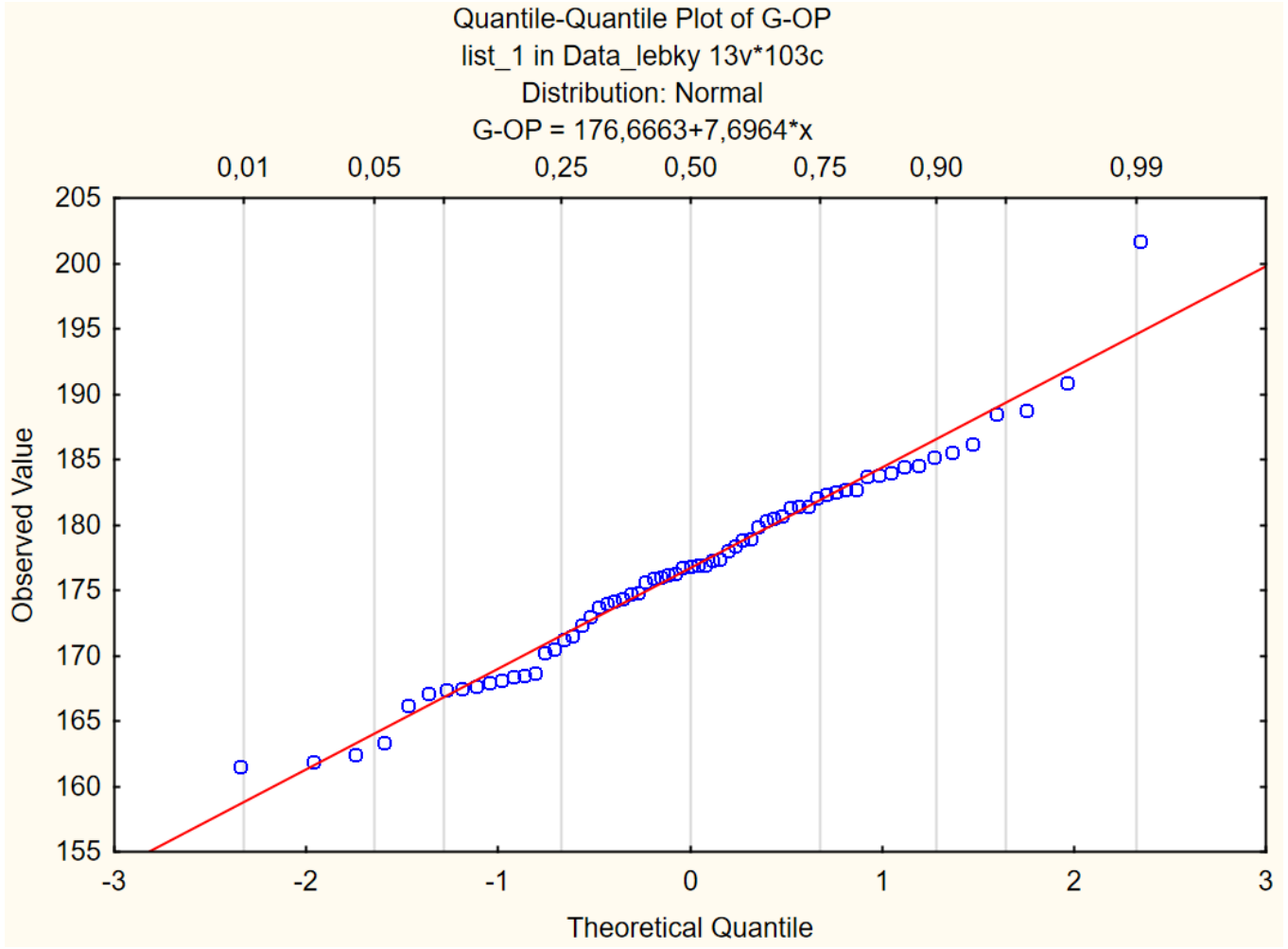




Umožňuje posoudit rozložení hodnot a srovnat je s předpokládaným rozložením (linie). Nastavuje se jako *Fit type* dialogovém okně.

Popisná statistika – vizuální hodnocení – QQ graf

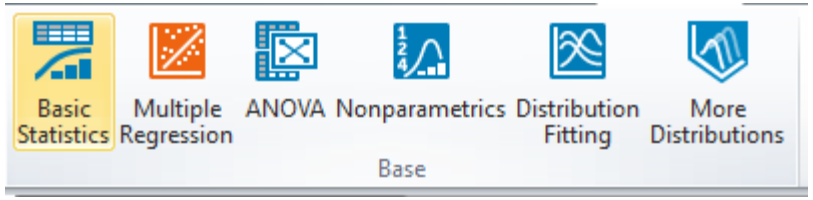


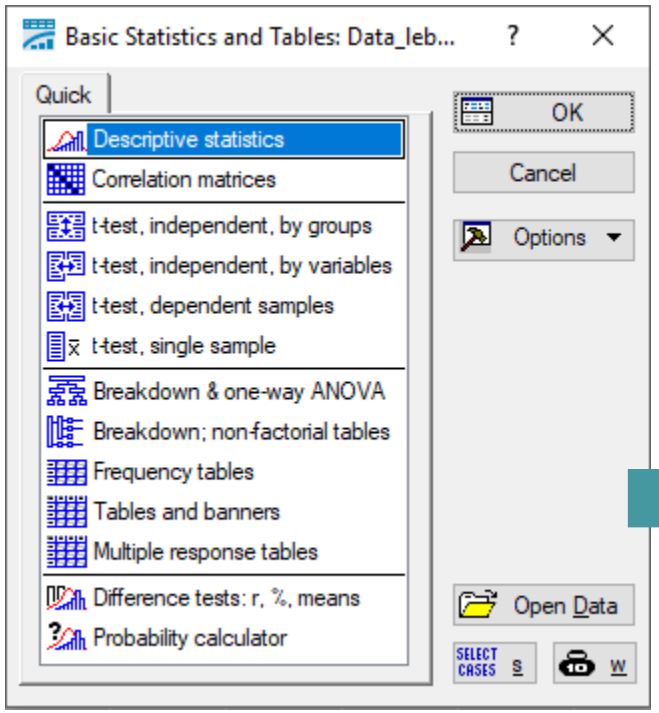


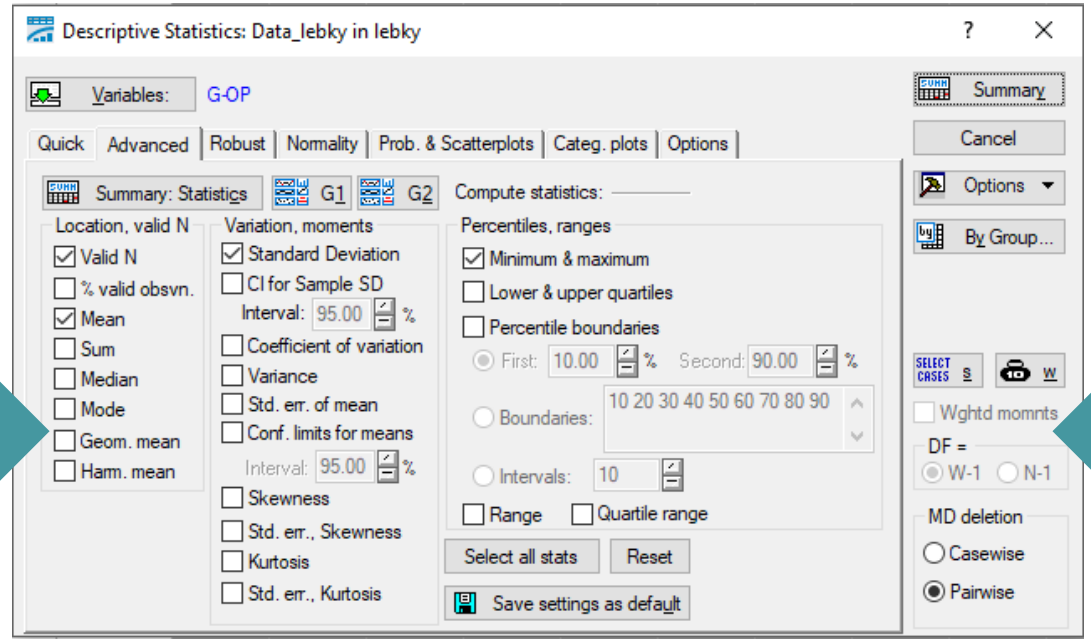
Alternativní způsob porovnání pozorovaných hodnot s normálním rozložením (pozorovaný kvantil vs. teoretický kvantil).

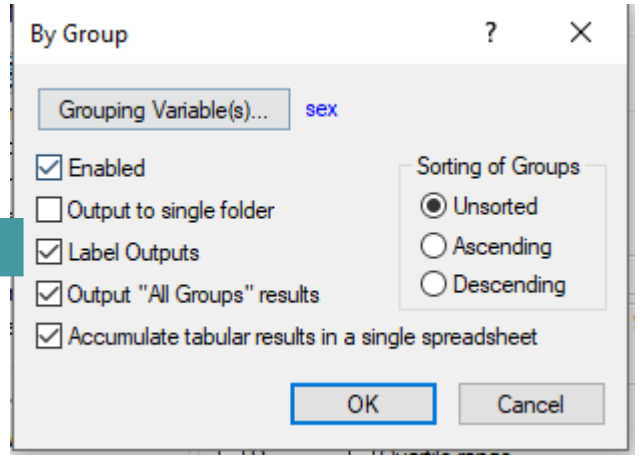
Popisná statistika – číselná popisná statistika

číselná popisná statistika









Variable	Descriptive Statistics (Data_lebky in lebky)						
	Valid N	Mean	Median	Minimum	Maximum	Variance	Std.Dev.
G-OP	65	176.6663	176.8000	161.5500	201.6700	58.86475	7.672336

pro všechna data zároveň

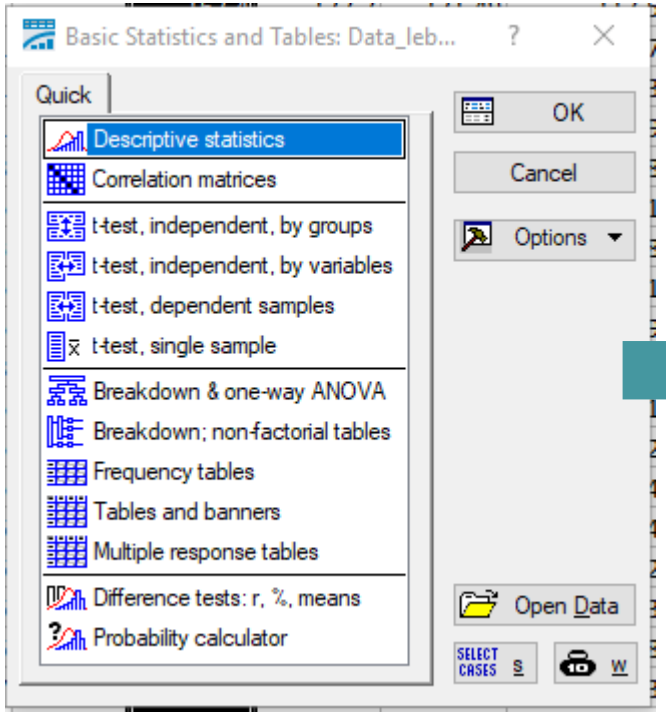
Aggregate Results
Descriptive Statistics (Spreadsheet in lebky)

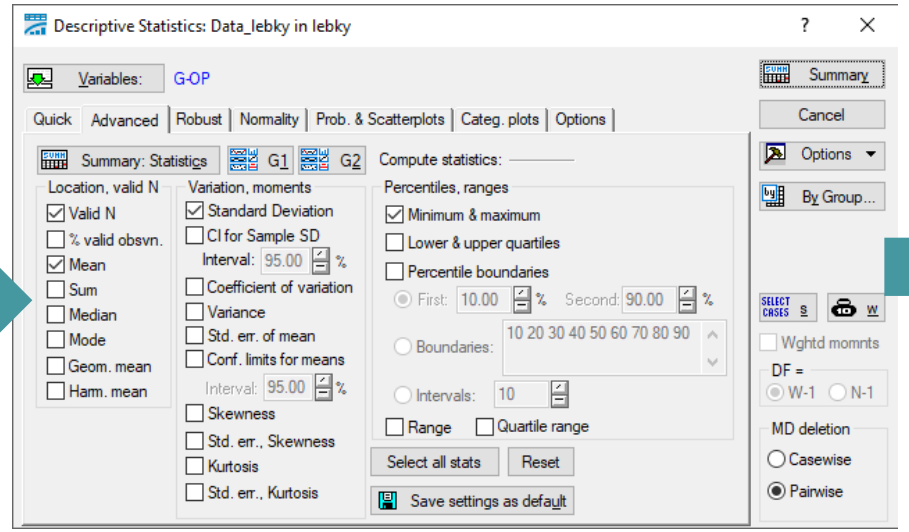
Variable	sex	Valid N	Mean	Median	Minimum	Maximum	Variance	Std.Dev.
G-OP	F	33	174.2264	175.6800	161.8900	184.5100	41.17876	6.417068
G-OP	M	32	179.1825	180.4300	161.5500	201.6700	66.14724	8.133095
G-OP		0						

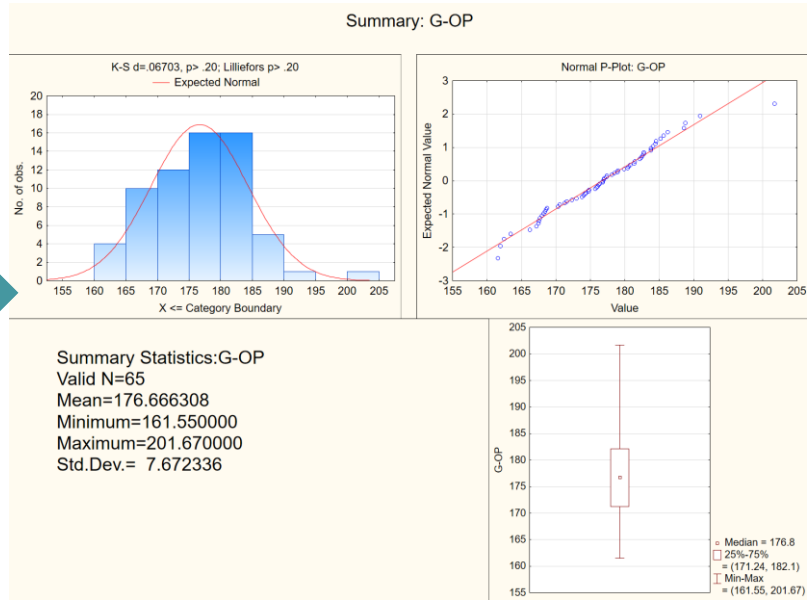
by group – pro skupiny zvlášť

Popisná statistika – číselná popisná statistika

souhrnné výsledky – histogram, krabicové grafy, zvolené parametry a P-P plot







v základní,
přednastavené podobě

Popisná statistika – normalita dat

Grafické posouzení

Srovnání s normálním rozložením – viz předchozí grafy

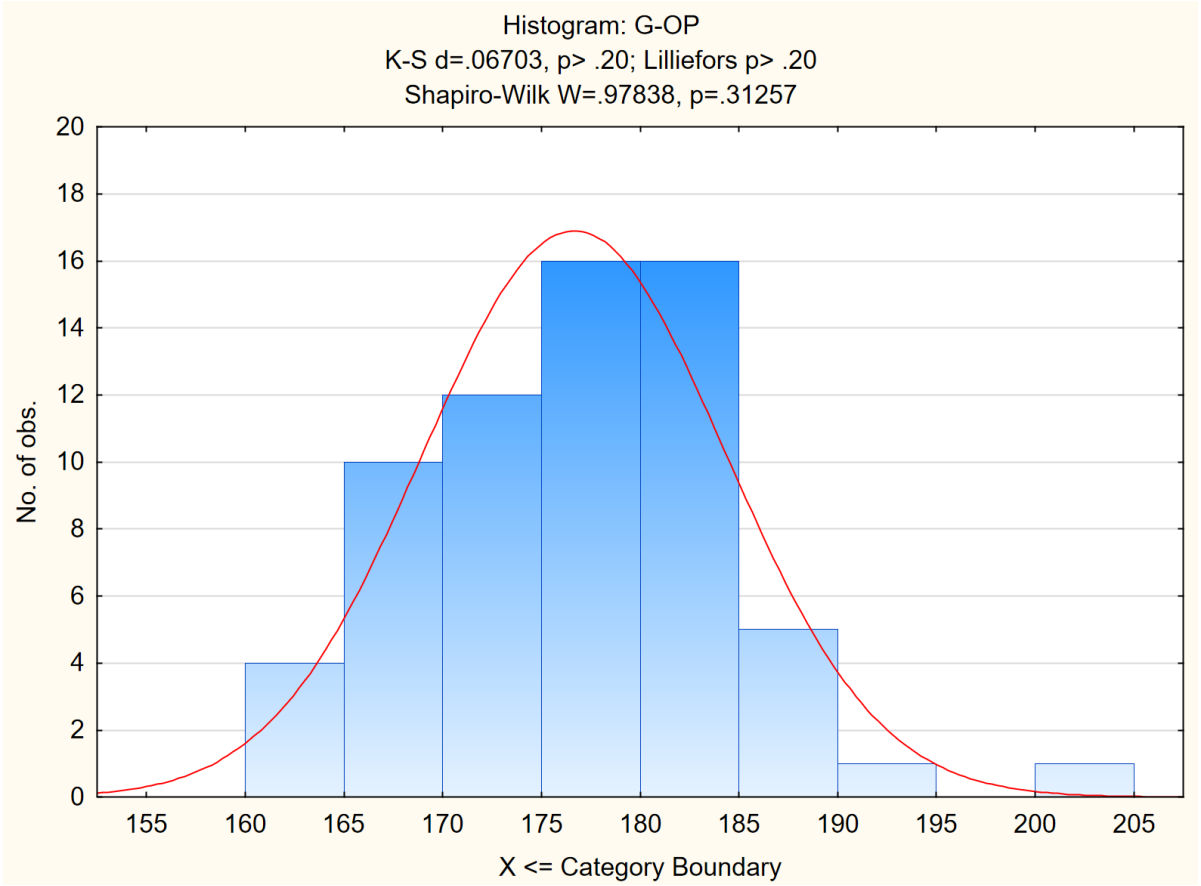
Testování

Statistické testy

Statistics > Basic statistics > Descriptive statistics > Normality

Frequency table: G-OP (Data_lebky in lebky)
K-S d=.06703, p> .20; Lilliefors p> .20

Category	Count	Cumulative Count	Percent of Valid	Cumul % of Valid	% of all Cases	Cumulative % of All
155.0000<x<=160.0000	0	0	0.00000	0.0000	0.00000	0.0000
160.0000<x<=165.0000	4	4	6.15385	6.1538	3.88350	3.8835
165.0000<x<=170.0000	10	14	15.38462	21.5385	9.70874	13.5922
170.0000<x<=175.0000	12	26	18.46154	40.0000	11.65049	25.2427
175.0000<x<=180.0000	16	42	24.61538	64.6154	15.53398	40.7767
180.0000<x<=185.0000	16	58	24.61538	89.2308	15.53398	56.3107
185.0000<x<=190.0000	5	63	7.69231	96.9231	4.85437	61.1650
190.0000<x<=195.0000	1	64	1.53846	98.4615	0.97087	62.1359
195.0000<x<=200.0000	0	64	0.00000	98.4615	0.00000	62.1359
200.0000<x<=205.0000	1	65	1.53846	100.0000	0.97087	63.1068
Missing	38	103	58.46154		36.89320	100.0000



T-test

Nepárový dvouvýběrový **t-test**

Řešená otázka

Je mezi dvěma skupinami v konkrétní kvantitativní proměnné významný rozdíl?

Jsou muži statisticky významně vyšší než ženy?

Párový dvouvýběrový **t-test**

Řešená otázka

Je mezi stejnými jedinci v různé situaci rozdíl?

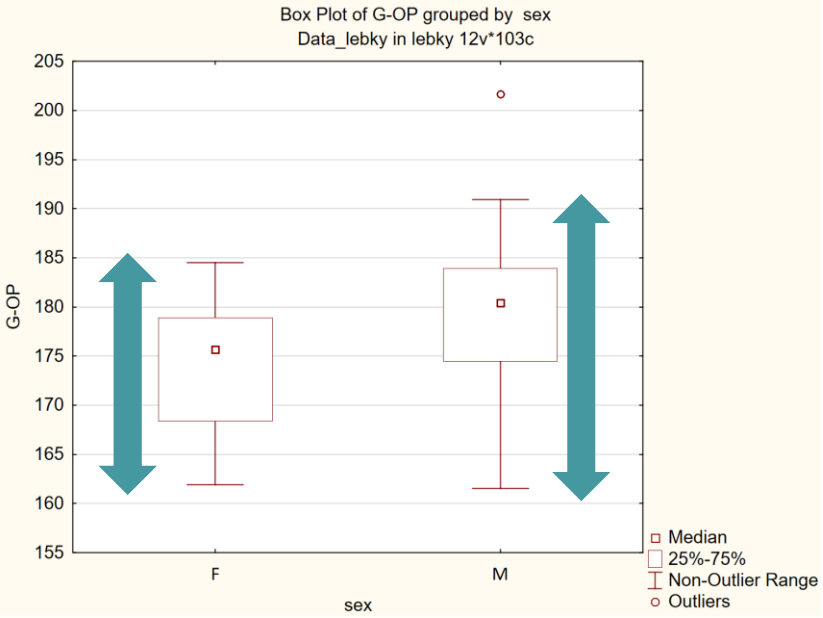
Jsou lidé po tréninku zdatnější než před ním?

T-test

Nepárový dvouvýběrový **t-test**

Předpoklady:
normální rozložení v rámci porovnávaných skupin
- **již představenými postupy**

shoda rozptylu těchto skupin
- testování je přímo součástí výsledků jako F-statistika



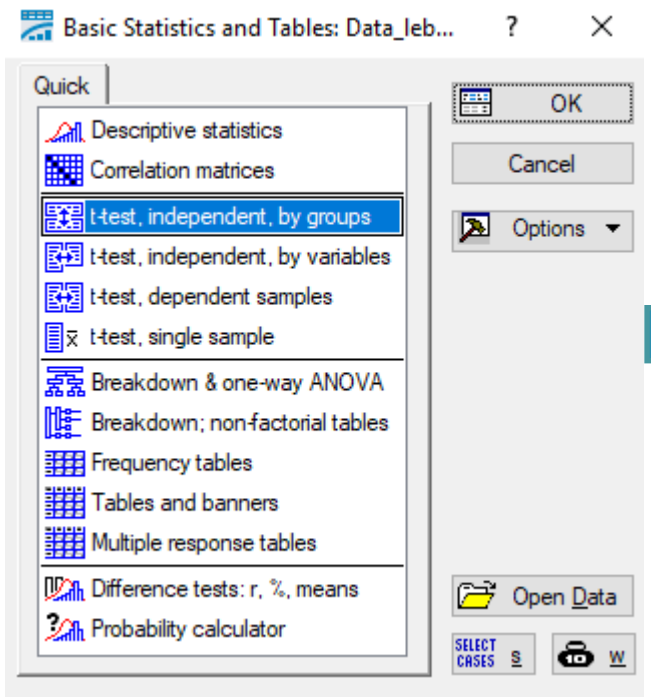
pokud data nesplňují

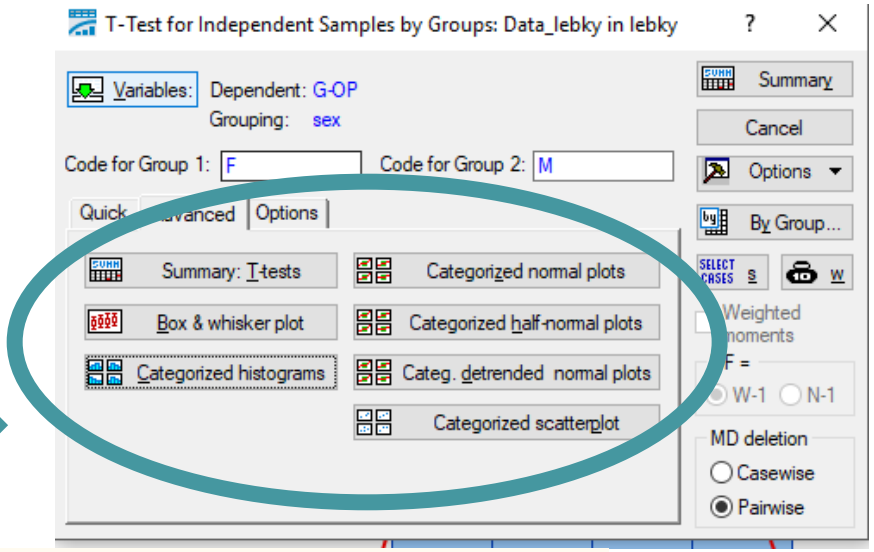
neparametrické alternativy
v tomto případě
Mann-whitney U-test

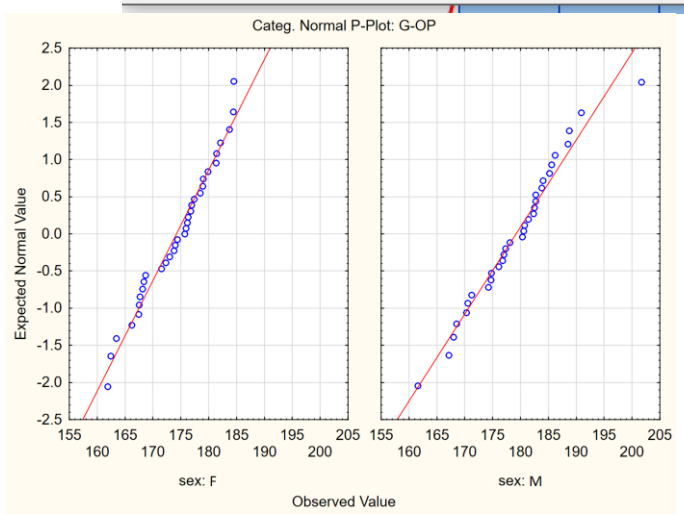
V případě různých rozptylů možno použít t-test se samostatnými odhady rozptylů

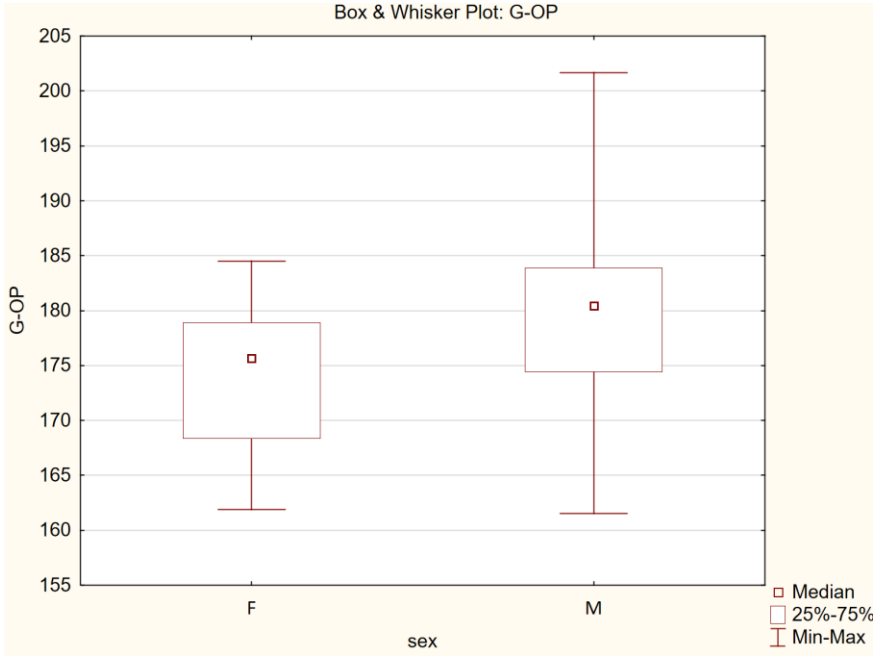
T-test

Ověření předpokladů testu přímo v dialogovém okně









Shoda rozptylů – graficky

Advanced > Box & Whiskers plot

Normalita rozložení v rámci skupin

Advanced > Categorized normal plots

T-test – výstupy

Samotné výstupy testu – lze provést hromadně pro všechny zároveň

T-tests; Grouping: sex (Data_lebky in lebky)
Group 1: F
Group 2: M

Variable	**Mean F**	Mean M	t-value	df	p	Valid N F	Valid N M	Std.Dev. F	Std.Dev. M	F-ratio Variances	p Variances
G-OP	174.2264	179.1825	-2.73202	63	0.008157	33	32	6.417068	8.133095	1.606344	0.187893
EU-EU	135.4297	141.4306	-3.37315	63	0.001275	33	32	5.635122	8.468520	2.258439	0.024809
BA-B	128.9491	132.4963	-2.51013	63	0.014650	33	32	4.961380	6.365777	1.646258	0.166251
ZYG-ZYG	120.5567	128.8028	-6.17618	63	0.000000	33	32	5.473546	5.284873	1.072676	0.846689
D-D	20.2794	21.4384	-1.89444	63	0.062759	33	32	2.369016	2.562275	1.169810	0.661057
RH-NS	33.0933	37.5522	-5.41930	63	0.000001	33	32	3.163414	3.467077	1.201199	0.608583
ZM-ZM	88.1052	92.2650	-2.89237	63	0.005243	33	32	6.449334	5.035607	1.640310	0.171598

skupinové průměry

samotná statistika

směrodatné odchylky

shoda rozptylů

Jde o jednorozměrný test, proto máme pro každou proměnnou samostatný řádek!

Diskriminační analýza

Jaké použít proměnné

význam mají pouze ty, které mají nějakou **souvislost s kategoriální proměnnou**

Hodnocení vztahu nezávislých proměnných a kategoriální proměnné

- t-test a ANOVA
- korelační analýza a XY grafy
- hlavní komponenty a faktorová analýza
- diskriminační analýza
- „expertní znalost proměnných“

redundantní proměnné **snižují stabilitu modelu** a mohou vést k nesmyslným výsledkům

korelační analýza

Diskriminační analýza

Vztah ke kategoriální proměnné

Samostatný **t-test** pro jednotlivé proměnné – pro dvě skupiny!!
(*Basic statistics > t-test, independent, by groups*)

ANOVA (*Basic statistics > Breakdows & One-way ANOVA; Analysis of variance*) – **pro dvě a více skupin** (pro dvě skupiny jsou výsledky obdobné jako t-test)

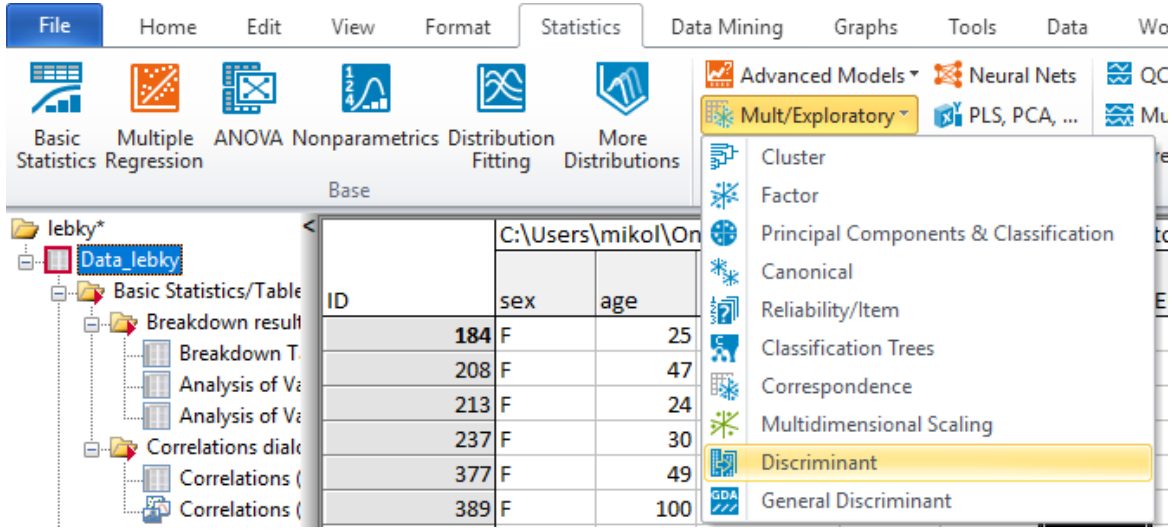
Analysis of Variance (Data_lebky in lebky)
Marked effects are significant at p < .05000

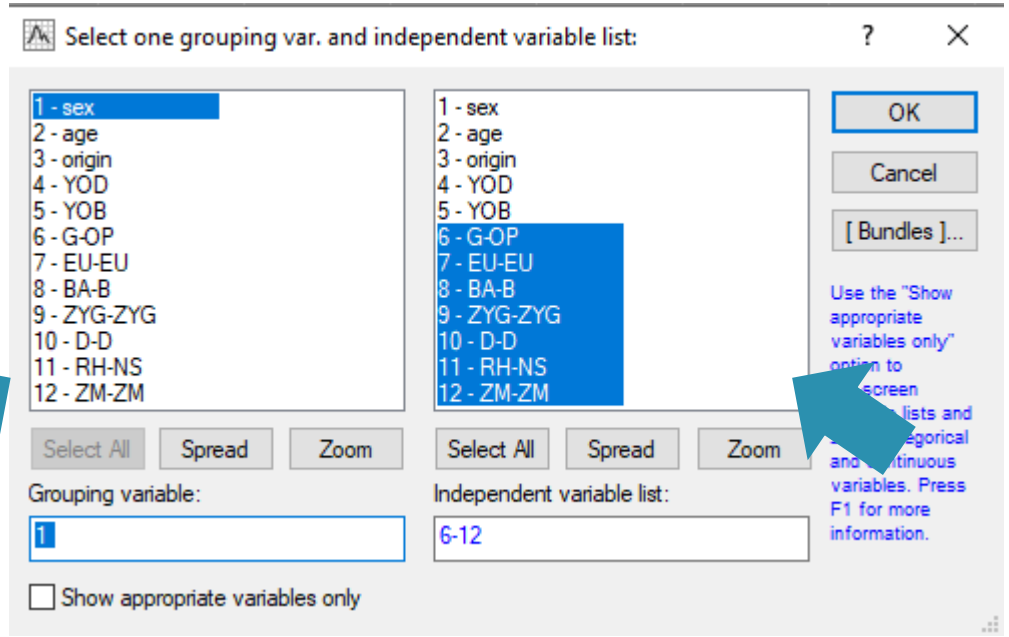
Variable	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
G-OP	399.059	1	399.059	3368.285	63	53.46484	7.46395	0.008157
EU-EU	585.042	1	585.042	3239.338	63	51.41806	11.37815	0.001275
BA-B	204.415	1	204.415	2043.906	63	32.44295	6.30074	0.014650
ZYG-ZYG	1104.721	1	1104.721	1824.537	63	28.96090	38.14525	0.000000
D-D	21.825	1	21.825	383.114	63	6.08118	3.58891	0.062759
RH-NS	322.996	1	322.996	692.869	63	10.99793	29.36881	0.000001
ZM-ZM	281.129	1	281.129	2117.083	63	33.60449	8.36582	0.005243

Obě analýzy mohou napovědět, ale diskriminace může uspět i díky kombinaci proměnných.

Diskriminační analýza

(*Statistics > Mult/Exploratory > Discriminant*)





grupovací proměnná – stav, který chceme určovat

nezávislá proměnná – výběr hodnot pro analýzu

Diskriminační analýza – číselný výstup analýzy

Celkové Wilks Lambda
celková kvalita modelu s použitím všech proměnných (0 = nejlepší diskriminace)

Wilks lambda celého modelu při vyřazení dané proměnné

Discriminant function Analysis Summary (Data_lebky in lebky)
No. of vars in model: 7; Grouping: sex (2 grps)
Wilks' Lambda: .51124 approx. F (7,57)=7.7849 p< .0000

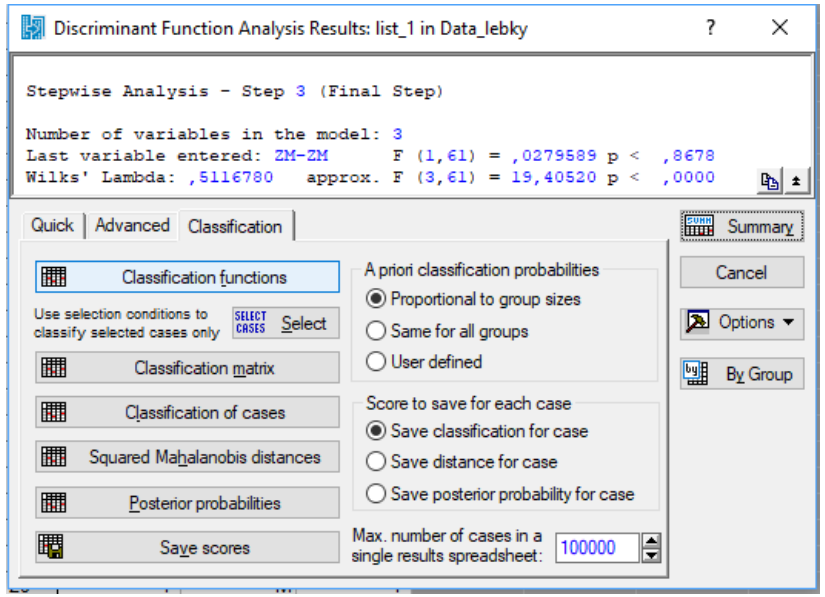
N=65	Wilks' Lambda	Partial Lambda	F-remove (1,57)	p-value	Toler.	1-Toler. (R-Sqr.)
G-OP	0.511237	1.000000	0.000000	1.000000	0.748612	0.251388
EU-EU	0.511532	0.999423	0.032915	0.856678	0.589176	0.410824
BA-B	0.511421	0.999640	0.020521	0.886598	0.800895	0.199105
ZYG-ZYG	0.564785	0.905188	5.970338	0.017668	0.493531	0.506469
D-D	0.511242	0.999989	0.000612	0.980357	0.825532	0.174468
RH-NS	0.600821	0.850898	9.988070	0.002523	0.882669	0.117332
ZM-ZM	0.527488	0.969192	1.811848	0.183616	0.852432	0.147568

Unikátní příspěvek dané proměnné k diskriminaci

Variabilita proměnné nevysvětlená ostatními proměnnými

Variabilita proměnné vysvětlená kombinací ostatních proměnných v modelu

Diskriminační analýza – predikční pravidlo



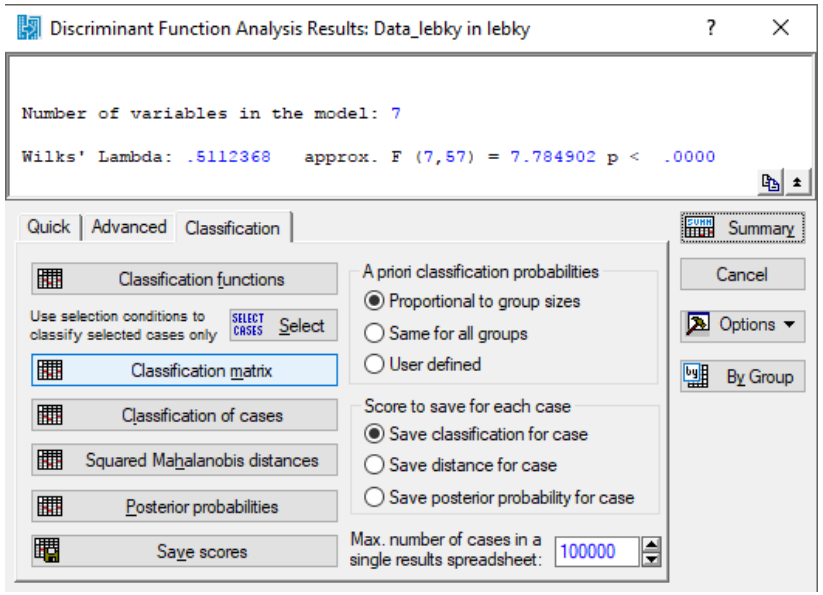
Classification Functions; grouping: sex (list_1 in Data_lebky)

Variable	F p=,50769	M p=,49231
ZYG-ZYG	3,362	3,577
RH-NS	2,157	2,503
ZM-ZM	1,972	2,065
Constant	-325,876	-373,301

rozepsané funkce pro jednu a pro druhou kategorii

spočítají se obě **a**
případ je klasifikován do té skupiny, pro kterou je výsledek vyšší

Diskriminační analýza – hodnocení klasifikačního kritéria



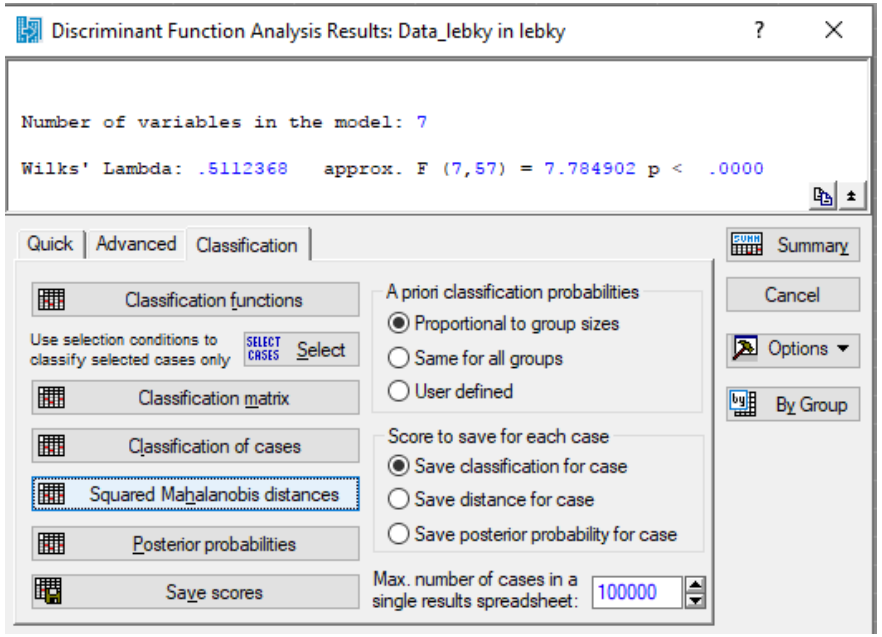
Classification Matrix (Data_lebky in lebky)
Rows: Observed classifications
Columns: Predicted classifications

Group	Percent Correct	F p=.50769	M p=.49231
F	84.84849	28	5
M	81.25000	6	26
Total	83.07692	34	31

klasifikační tabulka s procenty správně klasifikovaných případů (resubstituce)

Diskriminační analýza – další výstupy

Case	Observed Classif.	F p=.50769	M p=.49231
184	F	11.04751	17.18187
208	F	4.26211	7.58988
*213	F	6.96715	4.43218
237	F	3.90346	8.32681
377	F	1.85923	9.65731
389	F	4.18512	8.32306
475	F	4.78020	12.18686
491	F	8.18306	18.33574
5	M	4.54599	2.39293
*11	M	25.00021	26.27006
16	M	10.70175	5.22294
18	F	5.44417	15.17601
19	F	2.81365	4.45846
26	M	7.72133	6.64935
28	M	12.63816	7.46188



Case	Observed Classif.	F p=.50769	M p=.49231
	Posterior Probabilities (Data_lebky in lebk		
	Incorrect classifications are marked with *		
184	F	0.956808	0.043192
208	F	0.844836	0.155164
*213	F	0.225009	0.774991
237	F	0.903997	0.096003
377	F	0.980731	0.019269
389	F	0.890878	0.109122
475	F	0.976663	0.023337
491	F	0.993983	0.006017
5	M	0.260040	0.739960
*11	M	0.660540	0.339460
16	M	0.062466	0.937534
18	F	0.992584	0.007416
19	F	0.701234	0.298766
26	M	0.376316	0.623684
28	M	0.071933	0.928067
34	M	0.454583	0.545417

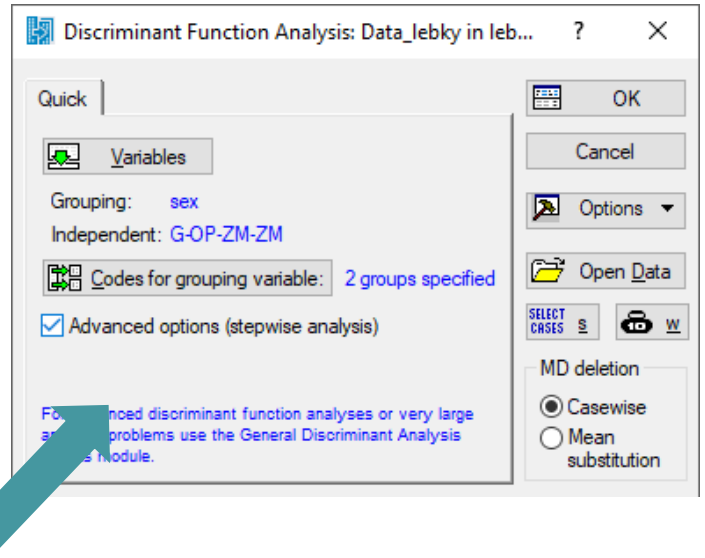
Mahalanobisova vzdálenost
– vzdálenost od centroidů obou skupin

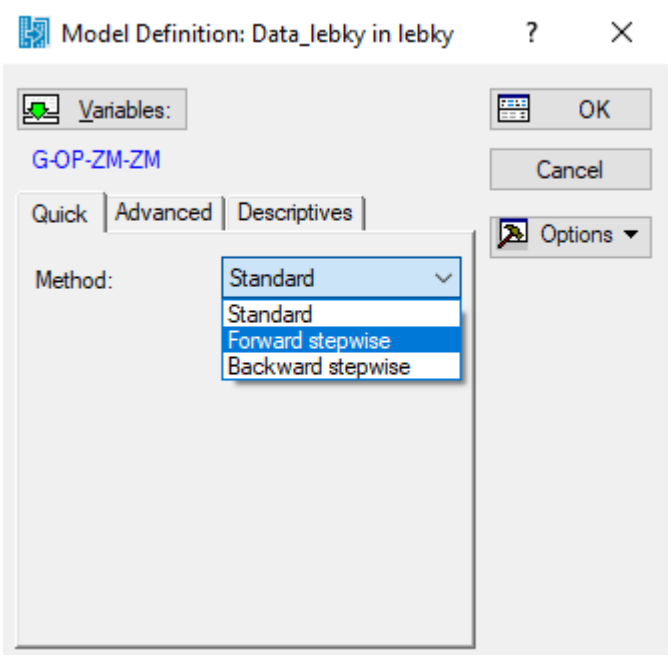
Aposteriorní pravděpodobnost
– pravděpodobnost, s jakou patří do obou skupin

Diskriminační analýza – dopředná a zpětná eliminace proměnných

„Step-wise“ analýza – výběr proměnných samotnou analýzou

- proměnné jsou přidávány/ubírány podle jejich významu v modelu
- zpravidla je vybrán pouze zlomek původních proměnných





Forward stepwise – dopředná
Backward stepwise – zpětná

Discriminant Function Analysis Summary (Data_lebky in lebky)
Step 3, N of vars in model: 3; Grouping: sex (2 grps)
Wilks' Lambda: .51168 approx. F (3,61)=19.405 p< .0000

N=65	**Wilks' Lambda**	Partial Lambda	F-remove (1,61)	p-value	Toler.	1-Toler. (R-Sqr.)
ZYG-ZYG	0.606466	0.843704	11.30024	0.001342	0.884970	0.115030
RH-NS	0.610534	0.838083	11.78516	0.001078	0.933334	0.066666
ZM-ZM	0.530334	0.964823	2.22404	0.141031	0.919489	0.080511

V tomto případě vybrány pouze tři proměnné

Kontingenční tabulky

Řešená otázka

Liší se nějak mezi skupiny jedinců ve výskytu znaků?

Test dobré shody
Testuje shodu reálné distribuce hodnot do n skupin s teoretickou distribucí

V **případě platnosti nulové hypotézy** je poměr mezi buňkami jednoho řádku v různých sloupcích nezávislý na výběru tohoto řádku

Statistics > Basic statistics > Tables and banners > Options > Expected frequencies

Advanced > Detailed Two-way Tables

pozorované

2-Way Summary Table: Observed Frequencies (Data_lebky in lebky
Marked cells have counts > 10

sex	origin admixed	origin European	Row Totals
F	8	25	33
M	0	32	32
Totals	8	57	65

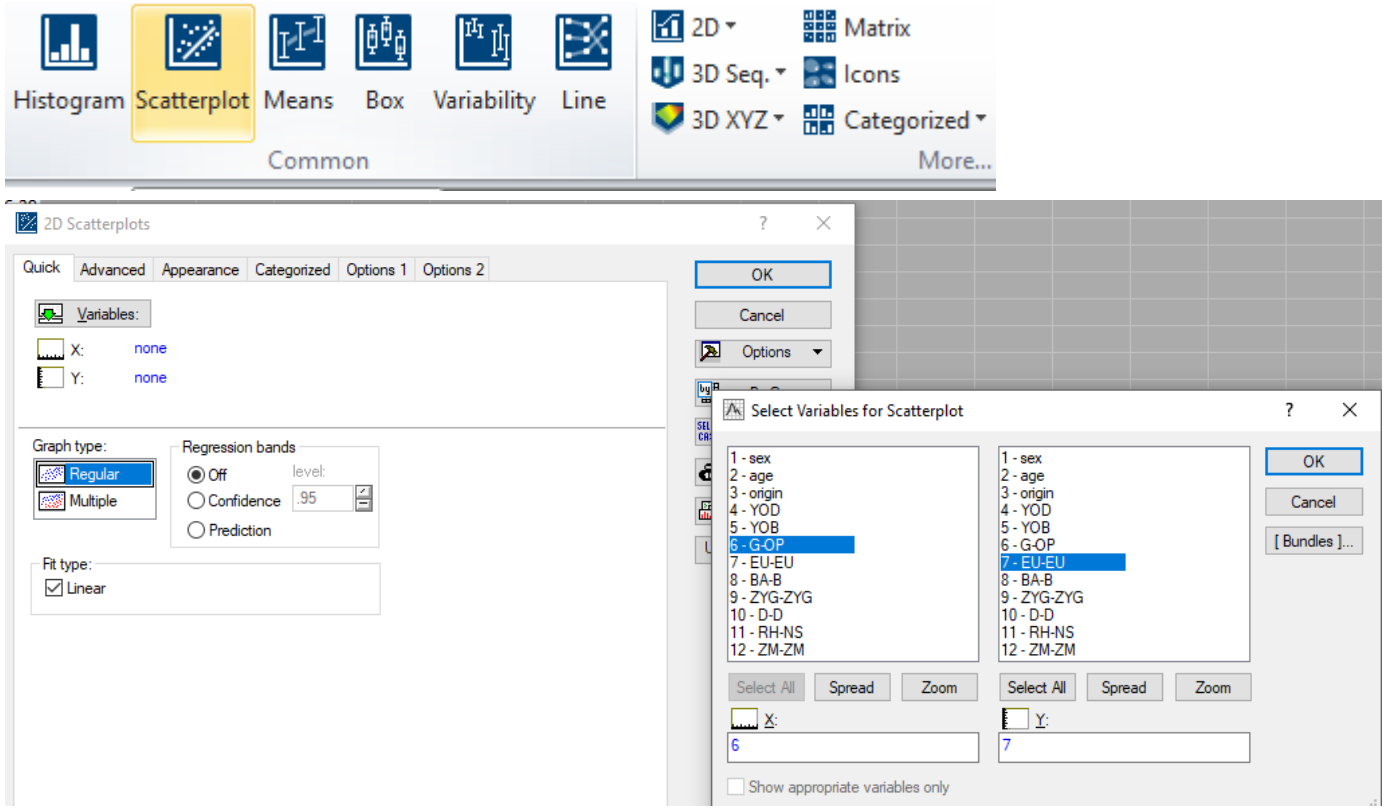
vs.

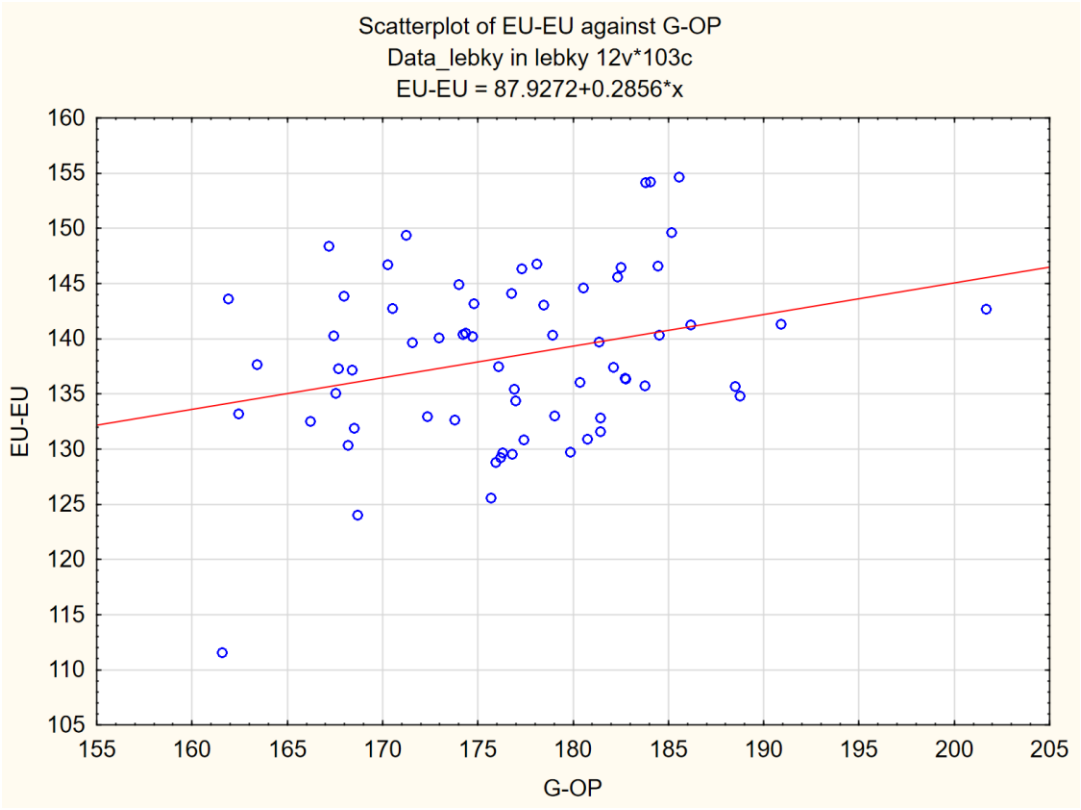
očekávané

2-Way Summary Table: Expected Frequencies (Data_lebky in lebky)
Marked cells have counts > 10

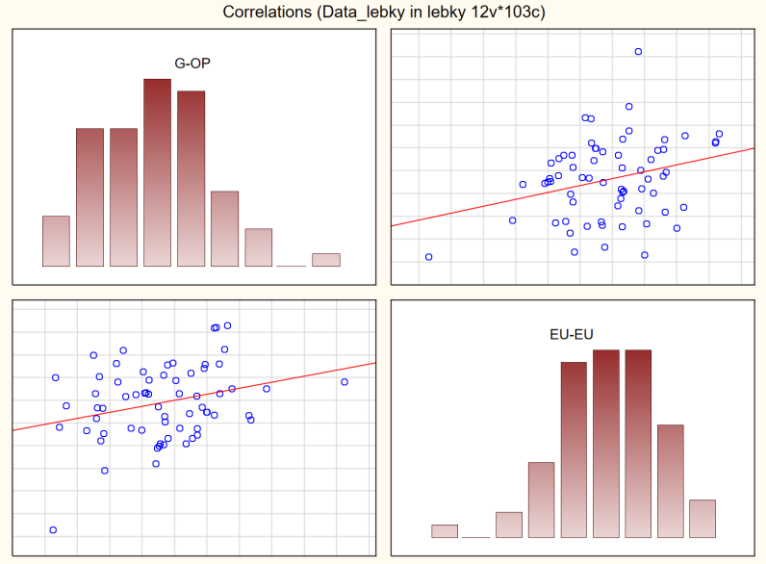
sex	**origin admixed**	origin European	Row Totals
F	4.061538	28.93846	33.00000
M	3.938462	28.06154	32.00000
Totals	8.000000	57.00000	65.00000

Korelační analýza – vizuální posouzení





Korelační analýza – číselné vyjádření – korelační koeficienty



Mají obě proměnné normální rozložení?

ANO?

Pearsonův korelační koeficient

Basic statistics > Correlation matrices

NE?

Spearmanův korelační koeficient

Non-parametrics > Correlations

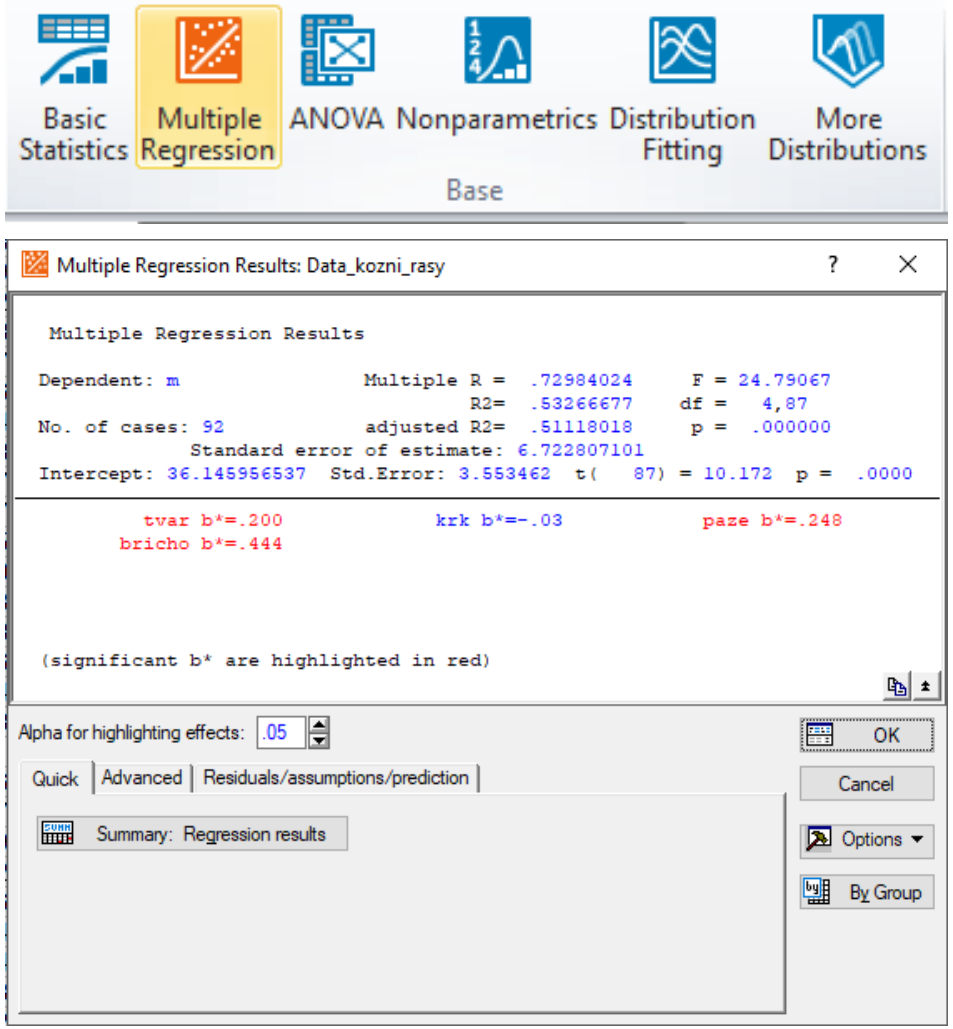
- pořadová korelace

Correlations (Data_lebky in lebky)
Marked correlations are significant at p < .05000
N=65 (Casewise deletion of missing data)

Variable	Means	Std.Dev.	G-OP	EU-EU	BA-B	ZYG-ZYG	D-D	RH-NS	ZM-ZM
G-OP	176.6663	7.672336	1.000000	0.283467	0.434146	0.449907	0.347321	0.277505	0.337226
EU-EU	138.3840	7.730197	0.283467	1.000000	0.312599	0.691211	0.242707	0.313714	0.172503
BA-B	130.6954	5.927057	0.434146	0.312599	1.000000	0.353512	0.245673	0.343633	0.215309
ZYG-ZYG	124.6163	6.765327	0.449907	0.691211	0.353512	1.000000	0.391823	0.485016	0.391089
D-D	20.8500	2.515388	0.347321	0.242707	0.245673	0.391823	1.000000	0.109204	0.292341
RH-NS	35.2885	3.984080	0.277505	0.313714	0.343633	0.485016	0.109204	1.000000	0.123358
ZM-ZM	90.1531	6.121443	0.337226	0.172503	0.215309	0.391089	0.292341	0.123358	1.000000

Vícenásobná regresní analýza - výstupy

$$Y = b_0 + b_1 X + E$$



Dependent – závislá (vysvětlovaná) proměnná
Multiple R – koeficient vícerozměrné korelace
R2 – koeficient determinace – **podíl modelem vysvětlované variability**
Adjusted R2 – podobný, ale bere v úvahu počet regresorů
F, df a p – F test vztahů mezi závislou proměnnou a množinou nezávislých proměnných
F = regresní průměr čtverců/reziduální průměr čtverců
Standard error of estimate – směrodatná chyba odhadu – rozptýlení pozorovaných hodnot kolem přímky
Intercept (Absolutní člen) – hodnota B0
Std. Error – směrodatná chyba absolutního členu (následují testy Ho – intercept je roven nule)
b* – standardizované koeficienty – umožňují porovnávat vliv jednotlivých proměnných

Regresní analýza

Další výsledky

Summary: regression results

První tabulka – statistiky z předchozího souhrnného okna

Druhá tabulka – podrobnější výsledky regrese, včetně nestandardizovaného koeficientu (b) (ten standardizovaný ukazuje relativní přispění jednotlivých proměnných)

Regression Summary for Dependent Variable: Váha (Data_somatometrie in
R= .65062825 R2= .42331712 Adjusted R2= .41619757
F(1,81)=59.458 p<.00000 Std.Error of estimate: 7.6022

N=83	b*	Std.Err. of b*	b	Std.Err. of b	t(81)	p-value
Intercept			-103.097	24.20592	-4.25917	0.000055
Výška	0.650628	0.084377	1.024	0.13275	7.71093	0.000000

Regression Summary for Dependent Variable: m (Data_kozni_rasy)
R= .72984024 R2= .53266677 Adjusted R2= .51118018
F(4,87)=24.791 p<.00000 Std.Error of estimate: 6.7228

N=92	b*	Std.Err. of b*	**b**	Std.Err. of b	t(87)	p-value
Intercept			36.14596	3.553462	10.17204	0.000000
tvar	0.200323	0.084188	1.38426	0.581752	2.37946	0.019524
krk	-0.026205	0.114921	-0.07819	0.342881	-0.22803	0.820159
paze	0.248340	0.110612	0.41047	0.182824	2.24515	0.027293
bricho	0.443919	0.108431	0.31736	0.077519	4.09402	0.000095

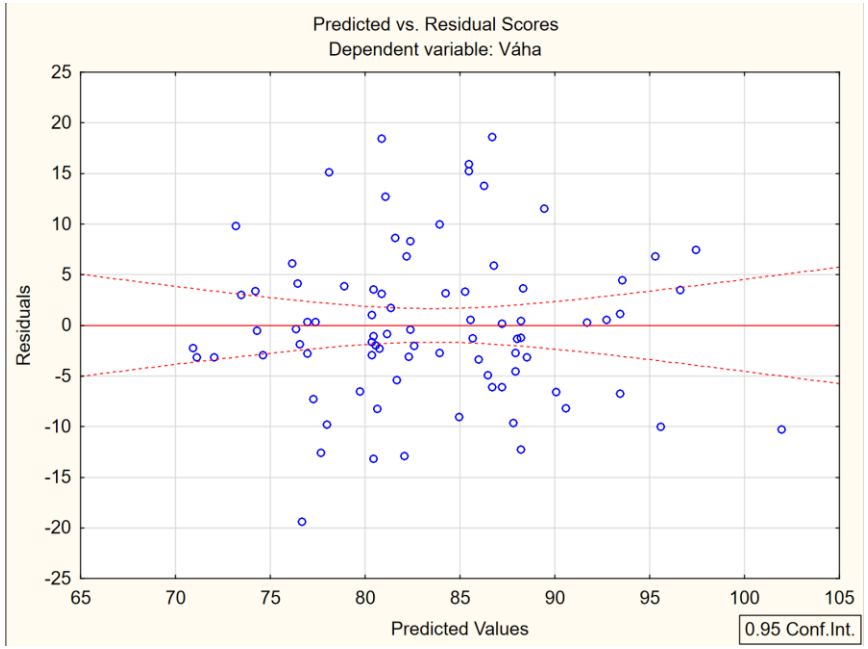
Pro každý koeficient jsou vypočítány hodnoty t-statistiky a p, které testují, zda je daný parametr významně odlišný od 0 (jestli má proměnná v modelu **své opodstatnění – součást verifikace modelu**).

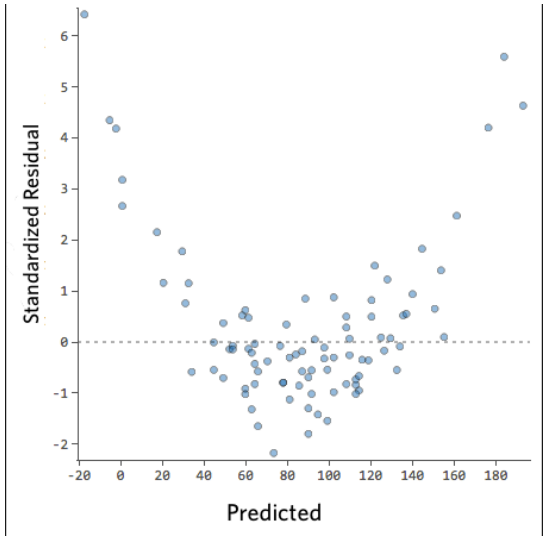
Regresní analýza

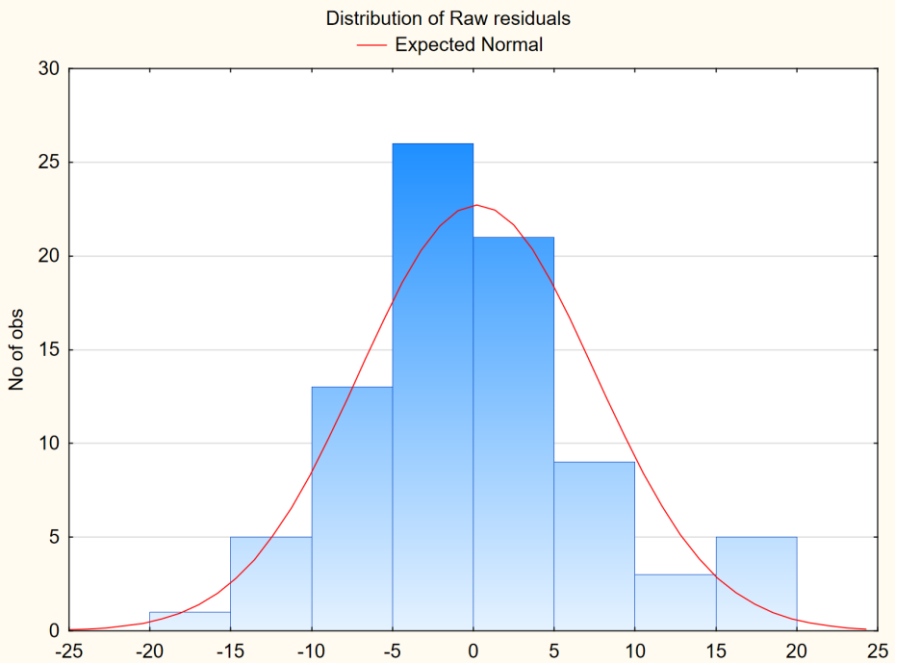
Ověření předpokladů

1) **Správně specifikovaný model**
2) **Střední hodnota chybové složky je 0**
3) **Chybová složka má konstantní rozptyl**
4) Jednotlivé složky chybového vektoru jsou nekorelova
5) Reziduální složka má normální rozdělení

Perform residual analysis > Scatterplots > Predicted vs. residuals







Perform residual analysis > Residuals > Histogram of residuals

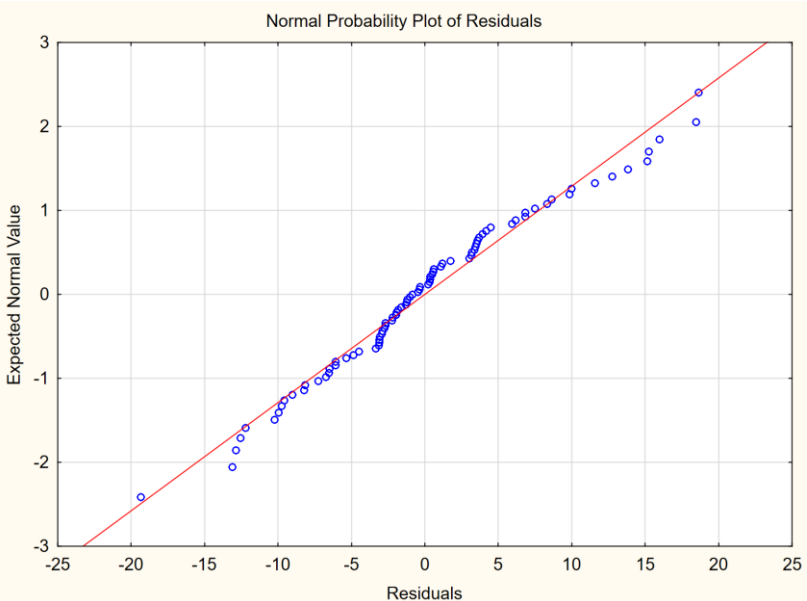
Regresní analýza – správná podoba výsledků

Ověření předpokladů

1) Správně specifikovaný model
2) Střední hodnota chybové složky je 0
3) Chybová složka má konstantní rozptyl
4) Jednotlivé složky chybového vektoru jsou nekorelované
5) **Reziduální složka má normální rozdělení**

máme nezávislé jedince

Perform residual analysis > Basics > Normal plot of residuals (Kvantilový graf)



V případě normality musí body ležet na proložené přímce.
Pokud neleží (dá se dále ověřit testem reziduí) – odhady parametrů modelu a regr. rovnice jsou v pořádku, ale **ne významnost regr. parametrů a konfidenční intervaly**

Regresní analýza – správná podoba výsledků

Predikce

Predict dependent variable

Compute confidence limits

Interval spolehlivosti pro průměrnou hodnotu odezvy

Udává rozmezí, kde se s 95 % spolehlivostí nachází *true best* fit populace

Compute prediction limits (interval předpovědi)

Interval spolehlivost pro individuální hodnotu odezvy

pokud použijete stejnou rovnici na další jedince dané výšky, bude se 95 % z nich nacházet v daném rozmezí