

Cvičení 4: Číselné charakteristiky dvourozměrného souboru, regresní přímka

Úkol 1.: Pro datový soubor ocel.sta vypočtete kovarianci a koeficient korelace mezi plasticity a meze pevnosti. Výsledek porovnejte s výsledkem ve skriptech Popisná statistika (str. 30).

Návod:

Výpočet kovariance: Statistika – Vícerozměrné průzkumné techniky – Hlavní komponenty&klasifikační analýza – Proměnné – Proměnné pro analýzu X, Y – OK – na záložce Details zaškrtneme Rozptyl počítat jako SČ/N – OK – na záložce Popisné statistiky vybereme Kovarianční matice.

Proměnná	Kovariance (ocel.sta)	
	X	Y
X	1052,403	985,763
Y	985,763	1057,207

Vysvětlení: Na hlavní diagonále jsou rozptyly proměnných X, Y, mimo hlavní diagonálu je kovariance.

Výpočet koeficientu korelace: Aktivujeme Výsledky hlavních komponent a klasifikační analýzy – na záložce Popisné statistiky vybereme Korelační matice.

Proměnná	Korelace (ocel.sta)	
	X	Y
X	1,000000	0,934548
Y	0,934548	1,000000

Vidíme, že výsledek 0,9345 svědčí o existenci silná přímé lineární závislosti mezi X a Y.

Úkol 2.: Pro údaje z datového souboru ocel.sta určete koeficienty regresní přímky meze pevnosti na mez plasticity a stanovte index determinace. Určete regresní odhad meze pevnosti, je-li mez plasticity 110. Nakreslete regresní přímku do dvourozměrného tečkového diagramu.

Návod: Statistika – Vícenásobná regrese – Závisle proměnná Y, nezávisle proměnná X - OK – OK – Výpočet: Výsledky regrese.

N=60	Výsledky regrese se závislou proměnnou : Y (ocel.sta) R= ,93454811 R2= ,87338017 Upravené R2= ,87119707 F(1,58)=400,06 p<0,0000 Směrod. chyba odhadu : 11,768					
	Beta	Sm.chyba beta	B	Sm.chyba B	t(58)	Úroveň p
Abs. člen			24,58814	4,740272	5,18707	0,000003
X	0,934548	0,046724	0,93668	0,046830	20,00160	0,000000

Ve výstupní tabulce najdeme koeficient b_0 ve sloupci B na řádku označeném Abs. člen, koeficient b_1 ve sloupci B na řádku označeném X, index determinace pod označením R2. Regresní přímka meze pevnosti na mez plasticity má rovnici:

$$Y = 24,58814 + 0,93668 X.$$

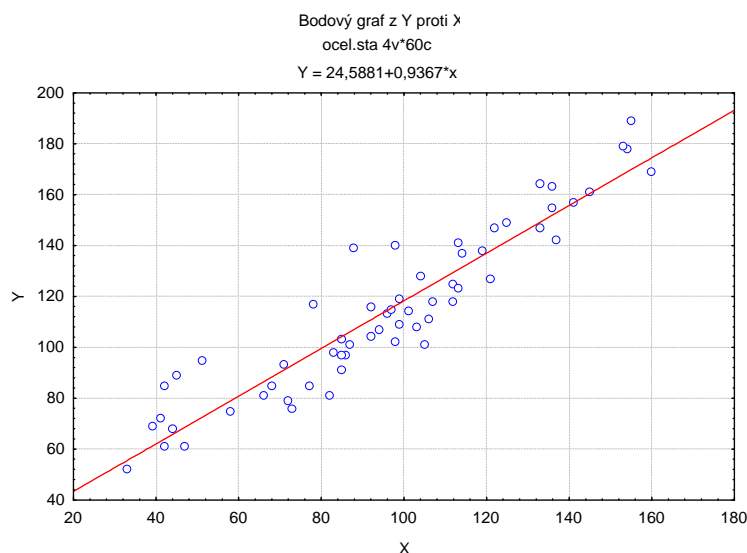
Index determinace nabývá hodnoty 0,8734, tedy variabilita meze pevnosti je z 87,34 % vysvětlena regresní přímkou.

Pro výpočet regresního odhadu zvolíme Rezidua/předpoklady/předpovědi - Předpověď závisle proměnné X:110 - OK. Ve výstupní tabulce je hledaná hodnota označena jako Předpověď.

Proměnná	Předpovězené hodnoty (ocel.sta) proměnné: Y		
	B-váž.	Hodnota	B-váž. * Hodnot
X	0,936679	110,0000	103,0346
Abs. člen			24,5881
Předpověď			127,6228
-95,0%LS			124,3063
+95,0%LS			130,9392

Je-li mez plasticity 110, je regresní odhad meze pevnosti roven 127,62.

Nakreslení regresní přímky: Do dvourozměrného tečkového diagramu nakreslíme regresní přímku tak, že v Bodovém grafu ponecháme Typ proložení Lineární, OK.



Na dvourozměrném tečkovém diagramu je vidět, že regresní přímka je vhodná pro modelování závislosti meze pevnosti na mezi plasticity – tečky jsou rozmístěny vcelku rovnoměrně kolem regresní přímky.

Úkol k samostatnému řešení

Máme k dispozici údaje o rozměrech lebek staroegyptské populace. Jedná se o 216 mužů a 109 žen.

Znak X ... největší délka mozkovny v mm (tj. přímá vzdálenost kranio-metrických bodů glabella a opisthocranion)

Znak Y ... největší šířka mozkovny v mm (tj. přímá vzdálenost kranio-metrických bodů euryon dx a euryon sin)

Znak Z ... pohlaví osoby (1 ... muž, 0 ... žena)

Údaje jsou uloženy v souboru lebky.sta.

- a) Určete koeficienty regresní přímky znaku Y na znak X, a to zvlášť pro muže a zvlášť pro ženy a stanovte index determinace.
 b) Určete regresní odhad znaku Y, nabývá-li znak X hodnoty 178 mm pro muže resp. 167 mm pro ženy.
 c) Nakreslete regresní přímku do dvourozměrného tečkového diagramu.

Výsledky pro muže:

Ad a)

Výsledky regrese se závislou proměnnou : Y (lebky.sta) R= ,16815700 R2= ,02827678 Upravené R2= ,02373602 F(1,214)=6,2273 p<,01333 Směrod. chyba odhadu : 4,7670 Zhrnout podmínku: z=1						
N=216	b*	Sm.chyba z b*	b	Sm.chyba z b	t(214)	p-hodn.
Abs.člen			114,0551	9,274551	12,29764	0,000000
X	0,168157	0,067385	0,1271	0,050919	2,49546	0,013335

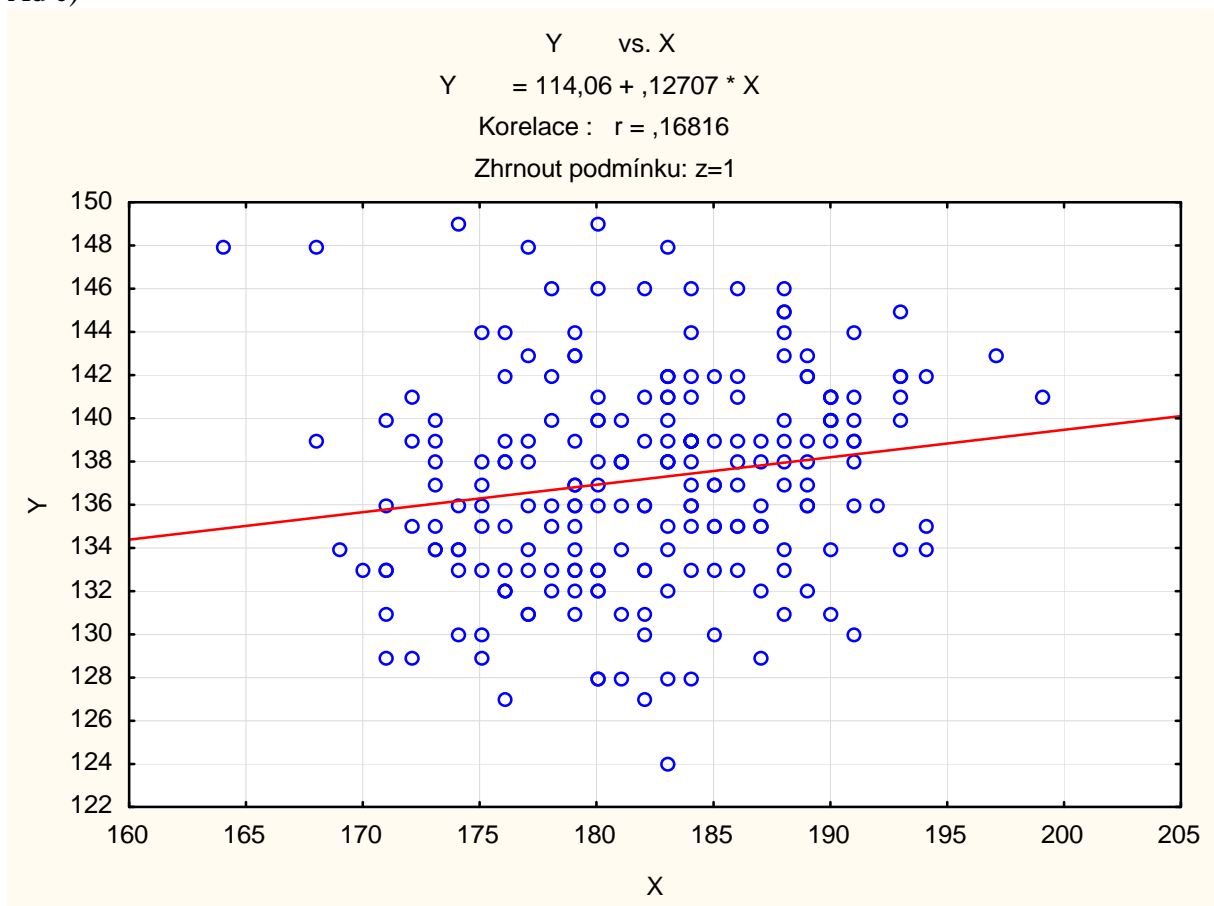
Regresní přímka má rovnici $Y = 114,0551 + 0,1271 X$.

Index determinace nabývá hodnoty 0,0283, tedy variabilita největší šířky mozkovny mužů je z 2,83 % vysvětlena regresní přímku.

Ad b)

Je-li největší délka mozkovny 178 mm, je regresní odhad největší šířky mozkovny roven 136,673 mm.

Ad c)



Výsledky pro ženy:

Ad a)

Výsledky regrese se závislou proměnnou : Y (lebky.sta)						
R= ,38092573 R2= ,14510442 Upravené R2= ,13711474						
F(1,107)=18,161 p<,00004 Směrod. chyba odhadu : 4,3622						
Zhrnout podmínku: z=0						
N=109	b*	Sm.chyba z b*	b	Sm.chyba z b	t(107)	p-hodn.
Abs.člen			83,94544	11,78726	7,121712	0,000000
X	0,380926	0,089385	0,28763	0,06749	4,261629	0,000044

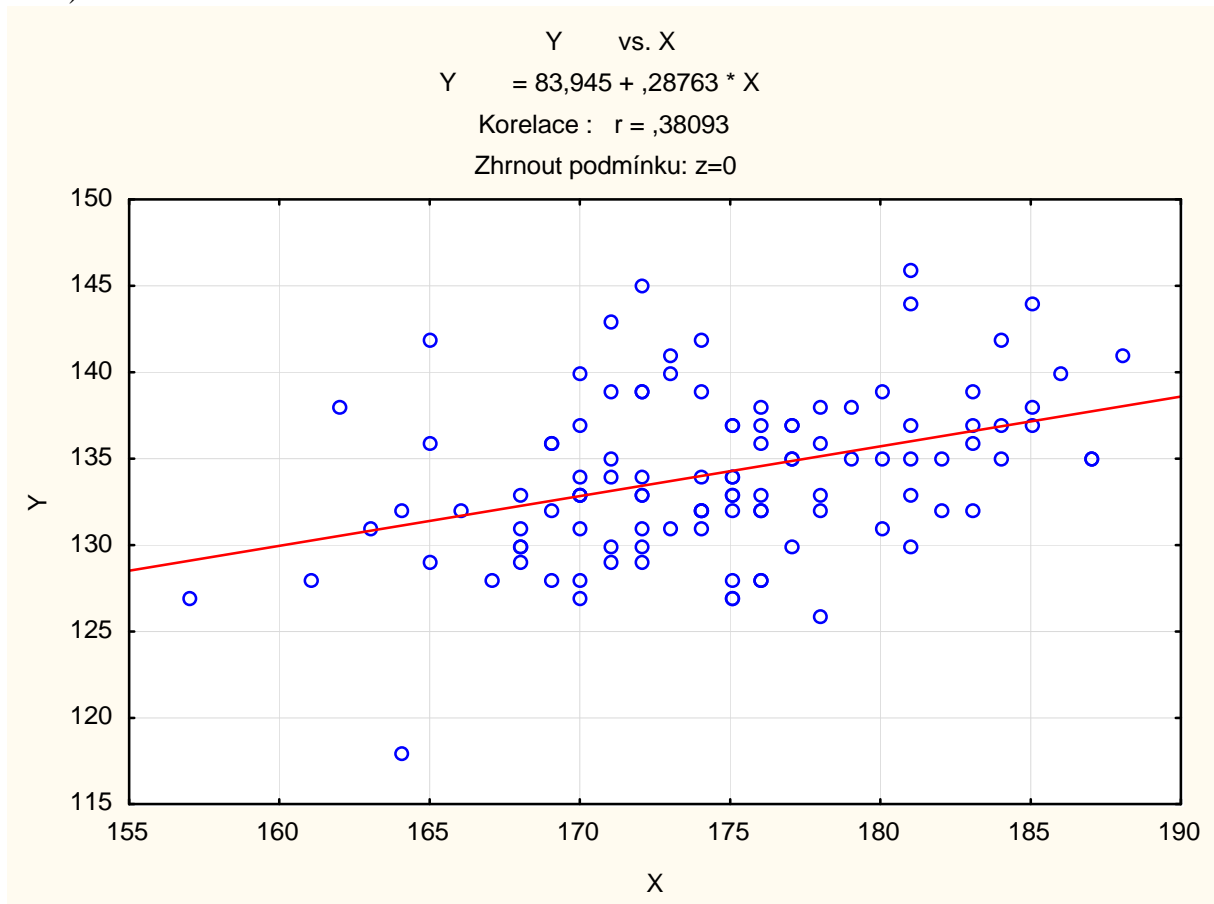
Regresní přímka má rovnici $Y = 83,9454 + 0,2873 X$.

Index determinace nabývá hodnoty 0,1451, tedy variabilita největší šířky mozkovny žen je ze 14,51 % vysvětlena regresní přímkou.

Ad b)

Je-li největší délka mozkovny 167 mm, je regresní odhad největší šířky mozkovny roven 131,98 mm.

Ad c)



Úkol k samostatnému řešení: V hodině tělesné výchovy bylo u 15 chlapců zjišťováno, kolik udělají shybů a kliků. Výsledky jsou uvedeny v tabulce.

číslo chlapce	X	Y
	počet shybů	počet kliků
1	1	10
2	3	15
3	2	15
4	0	0
5	5	40
6	6	25
7	1	7
8	4	31
9	3	30
10	5	35
11	6	41
12	2	10
13	1	14
14	1	9
15	8	64

Nakreslete dvourozměrný tečkový diagram vyjadřující závislost počtu kliků na počtu shybů.

Vypočtete koeficient korelace znaků X, Y.

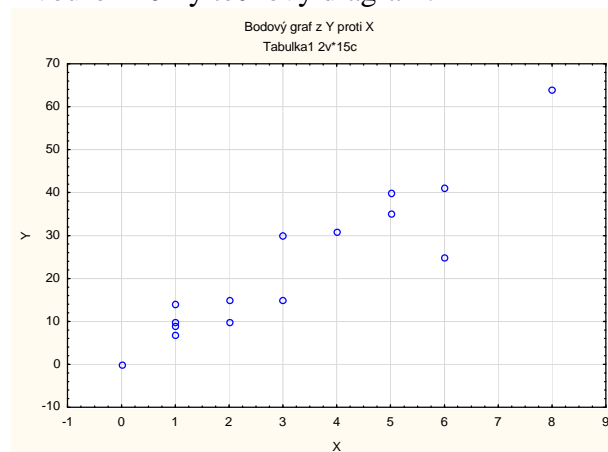
Modelujte závislost počtu kliků na počtu shybů regresní přímkou a napište její rovnici.

Vypočtete index determinace.

Jaký je regresní odhad počtu kliků pro počet shybů rovný 7?

Výsledek:

Dvourozměrný tečkový diagram:



Koeficient korelace nabývá hodnoty 0,9276.

Rovnice regresní přímkou: $Y = 1,6463 + 6,6939 X$.

Index determinace = 0,8604.

Regresní odhad počtu kliků pro 7 shybů je 48,5.