

## Vzorové řešení samostatné práce pomocí systému STATISTICA

**Úkol 1.:** V dílně pracuje 15 dělníků, u nichž byl zjištěn počet směn odpracovaných za měsíc (proměnná x) a počet zhotovených výrobků (proměnná y).

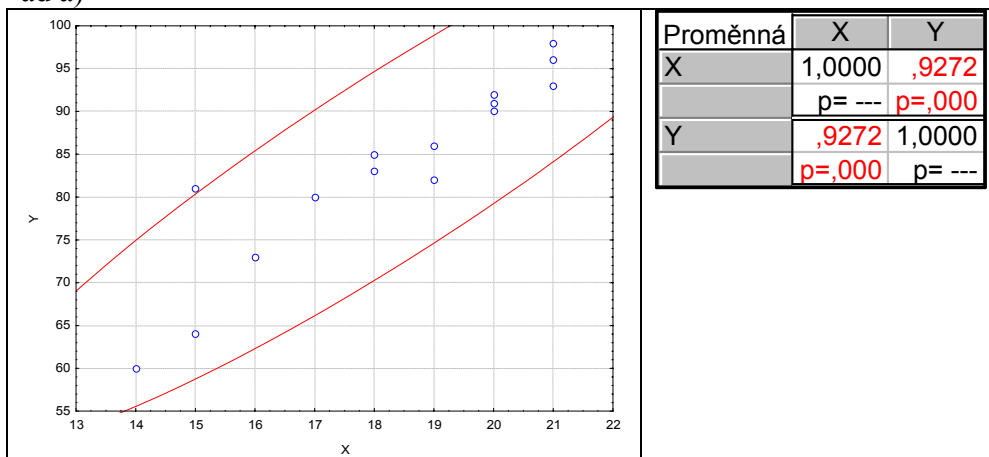
x: 20 21 18 17 20 18 19 21 20 14 16 19 21 15 15

y: 92 93 83 80 91 85 82 98 90 60 73 86 96 64 81

- Orientačně ověřte předpoklad, že data pocházejí z dvourozměrného normálního rozložení. Vypočtete výběrový koeficient korelace mezi x a y, interpretujte jeho hodnotu a na hladině významnosti 0,01 testujte hypotézu, že x a y jsou nezávislé náhodné veličiny.
- Najděte 99% asymptotický interval spolehlivosti pro skutečný koeficient korelace  $\rho$ . Za předpokladu, že regresní přímka dobře vystihuje závislost x na y
- vypočtete index determinace a interpretujte ho,
- na hladině významnosti 0,05 proveďte celkový F-test,
- najděte odhad rozptylu,
- odhadněte parametry regresní přímky, interpretujte směrnici regresní přímky a na hladině významnosti 0,05 proveďte dílčí t-testy,
- najděte 95% intervaly spolehlivosti pro regresní koeficienty,
- data s proloženou regresní přímkou a 95% intervaly spolehlivosti pro individuální predikované hodnoty a pro průměr predikovaných hodnot znázorněte graficky,
- jaká je predikovaná hodnota počtu výrobků pro 16 směn?
- zjistěte, zda rezidua se řídí normálním rozložením.

Řešení:

ad a)



ad b)

Dvourozměrný tečkový diagram svědčí o dvourozměrné normalitě dat. Výběrový koeficient korelace nabývá hodnoty blízké 1, tedy mezi počtem směn a počtem výrobků existuje velmi silná přímá lineární závislost – čím vyšší počet směn, tím vyšší počet výrobků. Protože p-hodnota je velmi blízká 0, hypotézu o nezávislosti zamítáme na hladině významnosti 0,01.  $0,714 < \rho < 0,983$  s pravděpodobností aspoň 0,99.

ad c)

$ID^2 = 0,8597$ . Znamená to, že regresní přímka závislosti počtu výrobků na počtu odpracovaných směn vystihuje variabilitu počtu výrobků téměř z 86%.

ad d)

Celkový F- test: testová statistika  $F = 79,63411$ , odpovídající p-hodnota = 0,000001. Na hladině významnosti 0,05 tedy zamítáme hypotézu, že dostačující je model konstanty.

ad e)

Odhad rozptylu:  $s^2 = 18,347$

ad f)

Výsledky regrese se závislou proměnnou : Y (Smeny.sta)						
R= ,92718009 R2= ,85966293 Upravené R2= ,84886777						
F(1,13)=79,634 p<,00000 Směrod. chyba odhadu : 4,2834						
N=15	Beta	Sm.chyba beta	B	Sm.chyba B	t(13)	Úroveň p
Abs.člen			5,010135	8,875949	0,564462	0,582049
X	0,927180	0,103900	4,302365	0,482123	8,923795	0,000001

$b_0 = 5,0101$ ,  $b_1 = 4,3024$ . Znamená to, že když se počet odpracovaných směn zvýší o 1, počet vyrobených výrobků se v průměru zvýší o 4,3.

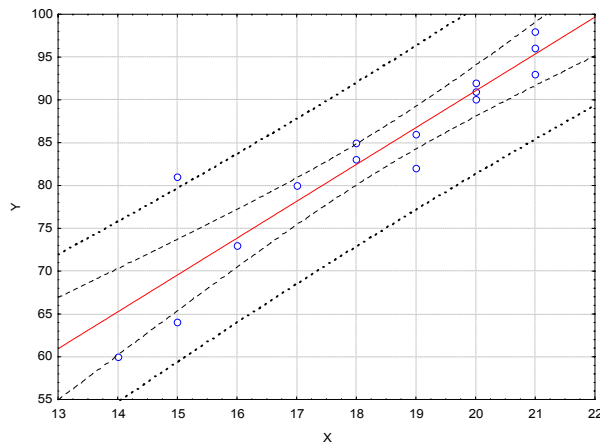
Dílčí t-testy: Na hladině významnosti 0,05 nezamítáme hypotézu, že regresní přímka prochází počátkem a zamítáme hypotézu, že počet výrobků nezávisí na počtu odpracovaných směn.

ad g)

$-14,1652 < \beta_0 < 24,1855$  s pravděpodobností aspoň 0,95

$3,2608 < \beta_1 < 5,3439$  s pravděpodobností aspoň 0,95

ad h)



ad i)

Predikovaná hodnota počtu výrobků pro 16 směn je 73,85.

ad j)

Normalita reziduí: Testová statistika K-S testu = 0,19453, Lilieforsova p-hodnota < 0,15, tedy hypotézu o normalitě reziduí nezamítáme na hladině významnosti 0,05.

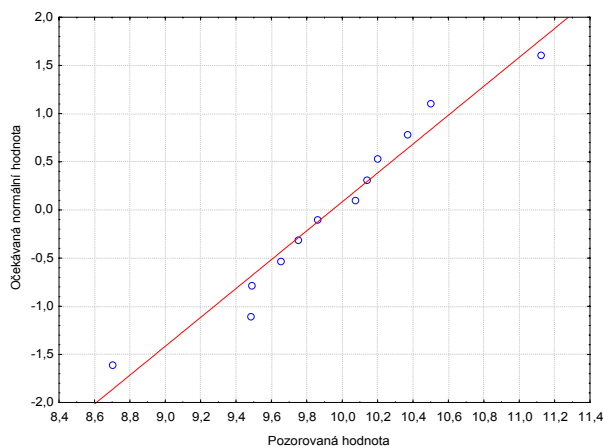
**Úkol 2.:** Uměle připravený vzorek minerálu obsahoval 10% křemene a byl dvanáctkrát proměřen.

Výsledky měření (v %): 8,7 10,2 10,07 9,75 9,65 10,37 10,14 10,5 9,48 11,22 9,49 9,86

- Vytvořte NP-plot a komentujte jeho vzhled.
- Na hladině významnosti 0,05 testujte pomocí K-S testu a S-W testu hypotézu, že uvedené údaje se řídí normálním rozložením.
- Na hladině významnosti 0,05 testujte hypotézu, že obsah křemene byl stanoven správně.

**Řešení:**

ad a)



Tečky s nevelkými odchylkami řadí vcelku uspokojivě podél přímky.

ad b)

Testová statistika K-S testu = 0,140003, Lilieforsova p-hodnota > 0,2.

Testová statistika S-W testu = 0,977, p-hodnota < 0,971.

Hypotézu o normalitě nezamítáme na hladině významnosti 0,05.

ad c)

Úloha vede na párový t-test. Testová statistika = -0,317, počet stupňů volnosti = 11, p-hodnota = 0,757. Na hladině významnosti 0,05 nelze zamítnout hypotézu, že obsah křemene byl stanoven správně.

**Úkol 3.:** Pro posouzení výkonnosti manuálního pracovníka během dne byl proveden experiment, při němž byla v různé denní dobu u sedmi pokusných osob měřena schopnost koncentrace, tj. počet správně provedených úkonů.

část dne	výkon						
ráno	162	162	150	151	164	155	155
dopoledne	158	154	150	160	165	156	149
odpoledne	149	150	146	158	154	152	148
večer	160	149	158	155	153	158	152
v noci	148	150	160	156	159	156	163

- Na hladině významnosti 0,05 testujte hypotézu, že denní doba nemá vliv na výkon.
- V případě zamítnutí hypotézy zjistěte, které výběry se liší.

c) Rozdílnost úrovně výkonu v různých denních dobách znázorníte graficky pomocí krabicového diagramu a komentujte jeho vzhled.

Řešení:

ad a)

Úloha vede na analýzu rozptylu jednoduchého třídění. Je zapotřebí ověřit, zda data v uvedených pěti skupinách se řídí normálním rozložením.

ID: 1 VYKON: SW-W = 0,8863; p = 0,2557; D = 0,2402; p < n.s.; Lil < 1

ID: 2 VYKON: SW-W = 0,9684; p = 0,8866; D = 0,1426; p < n.s.; Lil < 1

ID: 3 VYKON: SW-W = 0,9638; p = 0,8508; D = 0,1691; p < n.s.; Lil < 1

ID: 4 VYKON: SW-W = 0,9559; p = 0,7832; D = 0,2068; p < n.s.; Lil < 1

ID: 5 VYKON: SW-W = 0,9416; p = 0,6535; D = 0,2143; p < n.s.; Lil < 1

Ani S-W test ani K-S test nezamítá na hladině významnosti 0,05 hypotézu o normalitě.

Dále je zapotřebí ověřit předpoklad o shodě rozptylů.

Leveneův test homogenity rozptylů (Vykon.sta)								
Označ. efekty jsou význ. na hlad. p < ,05000								
Proměnná	SČ efekt	SV efekt	PČ efekt	SČ chyba	SV chyba	PČ chyba	F	p
VYKON	15,54286	4	3,885714	202,0000	30	6,733333	0,577086	0,681439

Protože p-hodnota Levenova testu je větší než hladina významnosti 0,05, hypotézu o shodě rozptylů nezamítáme.

Nyní provedeme analýzu rozptylu.

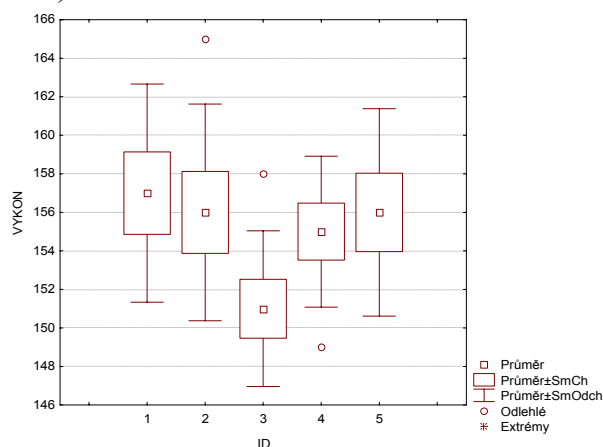
Analýza rozptylu (Vykon.sta)								
Označ. efekty jsou význ. na hlad. p < ,05000								
Proměnná	SČ efekt	SV efekt	PČ efekt	SČ chyba	SV chyba	PČ chyba	F	p
VYKON	154,0000	4	38,50000	746,0000	30	24,86667	1,548257	0,213652

Protože p-hodnota je větší než hladina významnosti 0,05, hypotézu o shodě středních hodnot nezamítáme. Znamená to, že na hladině významnosti 0,05 se neprokázal vliv denní doby na výkon manuálního pracovníka.

ad b)

Mnohonásobné porovnávání neprovádíme, protože se neprokázal vliv denní doby na výkon.

ad c)



Z diagramu je vidět, že variabilita výkonů se liší jen nepatrně. Nejhorší výkony podávají pracovníci odpoledne, nejlepší ráno.