

# POPISNÁ STATISTIKA

---

Komentované řešení pomocí programu *Statistica*

# Program Statistica I

- *Statistica* je velmi podobná *Excelu*.
- Na základní úrovni je to „klikací“ program určený ke statistickému zpracování dat.
- Data lze jednoduše vložit z *Excelu*
  - Otevřeme program *Statistica*. Zvolíme **Tabulka dat**. V okně **Vytvořit nový dokument** můžeme zvolit **Počet proměnných** (sloupců) **2** a **Počet případů** (řádků) **762** (jsou to rozměry našich dat) → **OK**
  - Metodou „Copy – Paste“ přeneseme data z *Excelu*

Vytvořit nový dokument

Tabulka dat

Počet proměnných: 2

Počet případů: 762

Délka jmen případů: 0

Kód ChD: -999999998

Výchozí datový typ: Double

Délka proměnné: 8

Umístění

v novém sešitě

jako samostatné okno

Prefix jmen prom: Prom

Počít číslo proměnné: A

Soubor Domů Upravit Zobrazit Formát Statistiky Data mining Grafy Nástroje Data

Vložit

Vymazat

Formát

Odstranit

Přesunout

Kopírování obrazovky

Nahradit

Opakovat

Najít

Přejít na

Náhodné hodnoty

dolů

doprava

Najít/nahradit

Vyplnit

	1 Prom1	2 Prom2			
1	43	65			
2	37	43			
3	27	7			
4	33	70			
5	41	19			
6	55	34			

# Program Statistica II

- Pojmenujeme si proměnné
  - Levým tlačítkem myši dvakrát klikneme na **Prom1**, objeví se okno, kde si můžeme nazvat první proměnnou

Dialogové okno "Proměnná 1" s nastaveními:

- Jméno: 1. PP
- Typ: Double
- Typ dat: Automaticky
- Délka: 8
- Kód ChD: -99999998
- Formát zobrazení: Obecné

Snímek obrazovky programu Statistica II s datovou tabulkou:

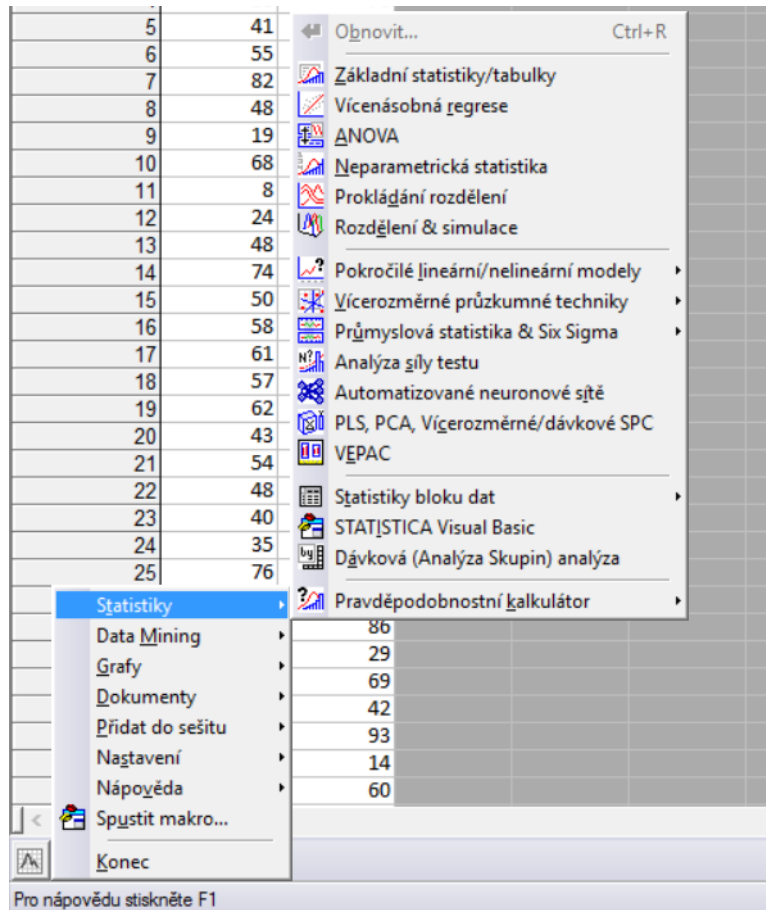
	1 1. PP	2 Prom2
1	43	65
2	37	43
3	27	7
4	33	70
5	41	19
6	55	34
7	82	50
8	48	52
9	19	11
10	68	88
11	8	3
12	24	16

- Stejně lze pojmenovat ostatní proměnné
- Výstupy z analýz lze ukládat do **Pracovního sešitu**, ze kterého se snadno přenášejí do vlastních prezentací

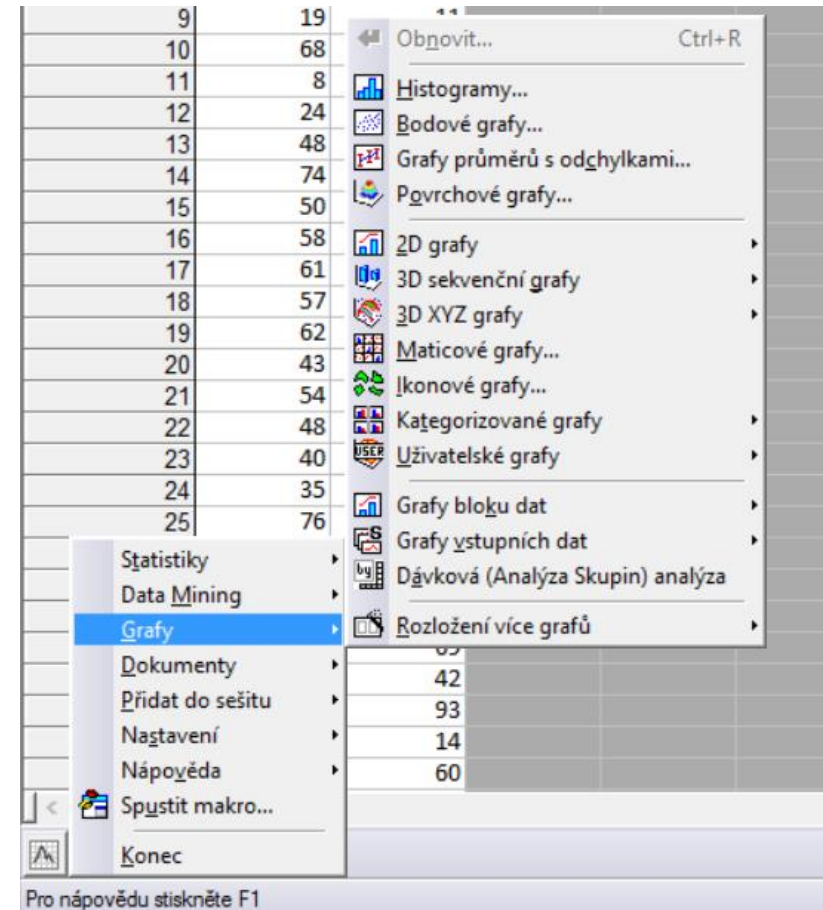
# Program Statistica III

- Základní analýzy lze rychle spouštět pomocí ikony umístěné vlevo dole

např.



či



# Program Statistica III

- Anebo nahore v listě, např.

The screenshot shows the Statistica software interface with the 'Statistika' ribbon selected. The ribbon contains various statistical analysis tools grouped into categories like 'Základ', 'Pokročilé/Vícerozměrné', 'Průmyslová statistika', and 'Nástroje'. The data table below the ribbon shows the following values:

	1	2
	1. PP	2. PP
1	43	65
2	37	43

či

The screenshot shows the Statistica software interface with the 'Grafy' ribbon selected. The ribbon contains various charting tools grouped into categories like 'Běžný', 'Více...', and 'Nástroje'. The data table below the ribbon shows the following values:

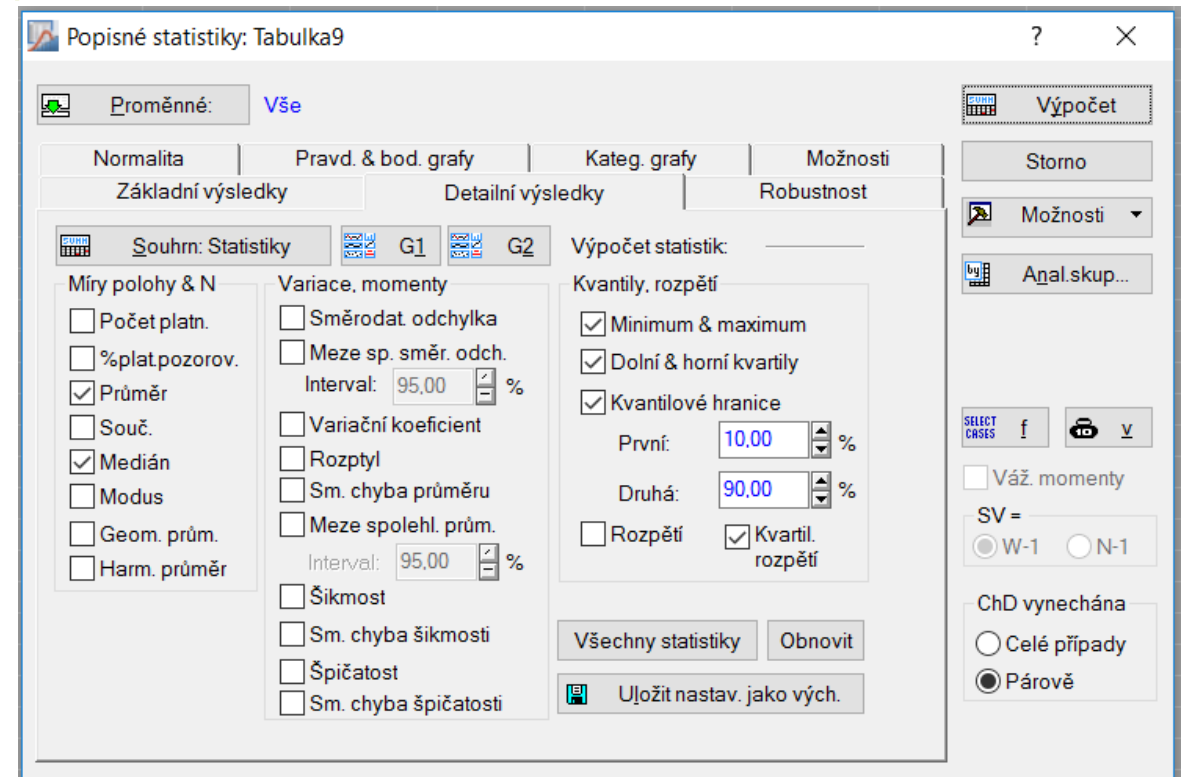
	1	2
	1. PP	2. PP
1	43	65
2	37	43

# Získaná data

- Máme k dispozici data o počtech bodů z prvního a druhého zápočtového testu z Matematiky I v zimním semestru 2015/2016, a to u všech 762 studentů, kteří psali oba testy.
- Zajímá nás, jak tyto písemky „dopadly“ a zdali výsledky z prvního a druhého zápočtového testu spolu nějak souvisí.

# Vybrané kvantily I

- Kvantily rozdělují uspořádaný soubor hodnot na dvě části dle předem zadaného poměru četností.
- **Statistiky** → **Základní statistiky/tabulky** → **Detailní výsledky**
  - Zvolíme si, jaké charakteristiky chceme spočítat
  - A dále **Proměnné** → **Vybrat vše** → **OK**
  - → **Výpočet**



# Vybrané kvantily II

- Získáme přehledné výsledky

Proměnná	Popisné statistiky								
	Průměr	Medián	Minimum	Maximum	Dolní kvartil	Horní kvartil	Kvantil 10,00000	Kvantil 90,00000	Kvartilové rozpětí
1. PP	46,57087	45,00000	0,00	100,0000	30,00000	65,00000	16,00000	79,00000	35,00000
2. PP	39,45013	39,00000	0,00	100,0000	21,00000	56,00000	8,00000	71,00000	35,00000

- Vidíme, že
  - výsledky z druhé písemky jsou o trochu horší než z první písemky.
  - Variabilita výsledků z obou písemek je však srovnatelná. Toto tvrzení bude zcela opodstatněné později na základě dalších charakteristik variability – zde máme zatím spočtené
    - **kvartilové rozpětí** – rozdíl horního a dolního kvartilu, které je u obou písemek stejné.



# Míry polohy

- Pro posouzení polohy (středního počtu bodů) použijeme dva ukazatele
  - **Aritmetický průměr** (patří do třídy momentových charakteristik)
  - **Medián** (kvantilová charakteristika)
- Postupem obdobným jako při hledání kvantilů, dostaneme:

Proměnná	Popisné statistiky	
	Průměr	Medián
1. PP	46,57087	45,00000
2. PP	39,45013	39,00000

- Na základě této analýzy (průměru i mediánu) vidíme, že výsledky z první zápočtové písemky jsou o něco lepší než z druhé písemky.
- Střední počet bodů u obou písemek je mírně pod padesát bodů.

# Míry variability I

- Variabilitu (proměnlivost, míru rozptýlení) posoudíme pomocí těchto ukazatelů:
  - **Rozptyl** (momentová charakteristika)
  - **Směrodatná odchylka** (momentová charakteristika)
  - **Variační rozpětí** (kvantilová charakteristika)
  - **Decilové rozpětí** (kvantilová charakteristika)
  - **Kvartilové rozpětí** (kvantilová charakteristika)
  - **Variační koeficient** (relativní míra variability)
- Stejným postupem získáme tabulku s výsledky
  - poslední sloupec – **Decilové rozpětí** – byl přidán do tabulky až nakonec a hodnoty dopočteny samostatně
  - název sloupce – **Rozpětí** – byl přejmenován dle naší terminologie na **Variační rozpětí**

# Míry variability II

Proměnná	Popisné statistiky									
	Průměr	Medián	Dolní kvartil	Horní kvartil	Variační rozpětí	Kvartilové rozpětí	Rozptyl	Sm.odch.	Var.koef.	Decilové rozpětí
1. PP	46,57087	45,00000	30,00000	65,00000	100,0000	35,00000	539,7092	23,23164	49,88450	63,00000
2. PP	39,45013	39,00000	21,00000	56,00000	100,0000	35,00000	553,6618	23,53002	59,64497	63,00000

- Momentové i kvartilové charakteristiky ukazují, že (absolutní) variabilita počtu bodů v obou testech je velmi podobná.
- V průměru se počet dosažených bodů od celkového průměru liší přibližně o 23 bodů (směrodatná odchylka).
- Bodové výsledky 80 % studentů se nachází v intervalu šířky 63 bodů (decilové rozpětí), počty bodů 50 % studentů jsou koncentrovány v intervalu šířky 35 bodů (kvartilové rozpětí).
- Relativní variabilita (směrodatná odchylka vztažena ku průměrnému počtu bodů) je u prvního testu trochu vyšší než u druhého testu, a to kvůli vyššímu průměru.

# Míra (lineární) závislosti I

- Sílu lineární závislosti posoudíme pomocí **Pearsonova korelačního koeficientu**, který je odvozen od druhých momentů obou proměnných
- Zvolíme **Statistiky** → **Základní statistiky/tabulky** → **Korelační matice** → **1 seznam proměn.** → **Vybrat vše** → **OK** → **Výpočet**

Proměnná	Korelace			
	Průměry	Sm.odch.	1. PP	2. PP
1. PP	46,57087	23,23164	1,000000	0,678991
2. PP	39,45013	23,53002	0,678991	1,000000

- Korelační koeficient je **0,679**.
  - Je označen červeně, protože je významně nenulový (na hladině významnosti  $\leq 5\%$ ).
  - Kladná hodnota korelačního koeficientu svědčí o pozitivní lineární závislosti mezi počty bodů z prvního a druhého testu. Měl-li student nadprůměrný počet bodů v prvním testu, dá se očekávat, že měl i nadprůměrný počet bodů v druhém testu (a naopak).

# Míra (lineární) závislosti II

- Nelze ovšem říci, že studenti se ve druhém testu nezhoršovali. Vzhledem k tomu, že průměrný počet bodů ve druhém testu je nižší než v prvním, mohli se studenti zhoršit i v případě pozitivní závislosti. Ta se totiž týká **porovnání počtu bodů s průměrem!**

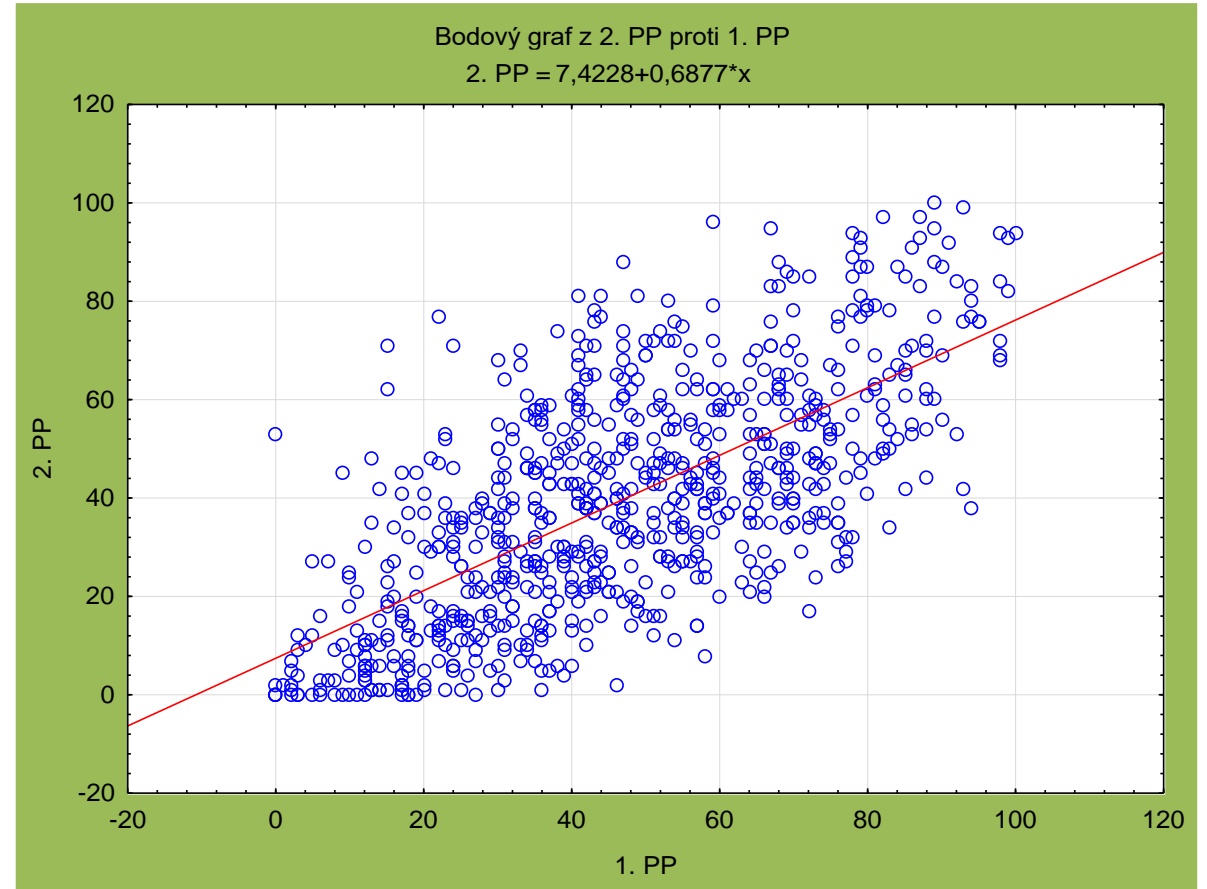
Posuzování intenzity závislosti podle velikosti korelačního koeficientu je arbitrární a v různých úlohách se může lišit. Často se však užívá obecné, empirické, pravidlo (viz tabulka vedle). Podle tohoto pravidla svědčí hodnota korelačního koeficientu 0,68 o relativně silné pozitivní závislosti, ne však perfektní závislosti. Znamená to tedy, že v některých případech se mohli studenti zhoršit i v porovnání s ostatními studenty (přesněji v porovnání s průměrem). Těchto případů však nebude mnoho.

Korel. koef.	Interpretace
-1	perfektní negativní lineární závislost
-0,7	silná negativní lineární závislost
-0,5	mírná negativní lineární závislost
-0,3	slabá negativní lineární závislost
0	žádná lineární závislost
0,3	slabá pozitivní lineární závislost
0,5	mírná pozitivní lineární závislost
0,7	silná pozitivní lineární závislost
1	perfektní pozitivní lineární závislost

# Bodový graf

- Sílu i tvar závislosti dobře ilustruje bodový graf

- Grafy → Bodové grafy → Proměnné – na osu x zvolíme 1. PP a na osu y 2. PP → OK → OK
- Bodový graf potvrzuje existenci relativně silné pozitivní lineární závislosti mezi počty bodů z prvního a druhého testu. Závislost však není perfektní – někteří studenti se dosti zlepšili ve druhém testu (body vlevo nahoře), někteří zase zhoršili (body vlevo dole). Většina studentů je však soustředěna kolem regresní přímky.



# Krabicový graf (box plot) I

- Grafy → Krabice
  - nastavení

2D krabicové grafy

Základní | Detaily | Vzhled | Kategorizovaný | Možnosti 1 | Možnosti 2

Typ grafu:

- Krabicové
- Běžný
- Max-Min, uzavírací
- Vícenásobný

Proměnné:  Závislé prom.: Vše  Grupovací prom.: žádné

Grupovací intervaly

- Celočís. mód  Auto
- Všechny hodnoty
- Neseřř.  123  321
- Kategorie: 10
- Kódy: žádné

Max (neodlehle)

75%

Medián

25%

Min (neodlehle)

Sřední bod

Hodn.: Medián

Styl: Čára

Společný rozptyl

OK

Storno

Možnosti

Anal. skup.

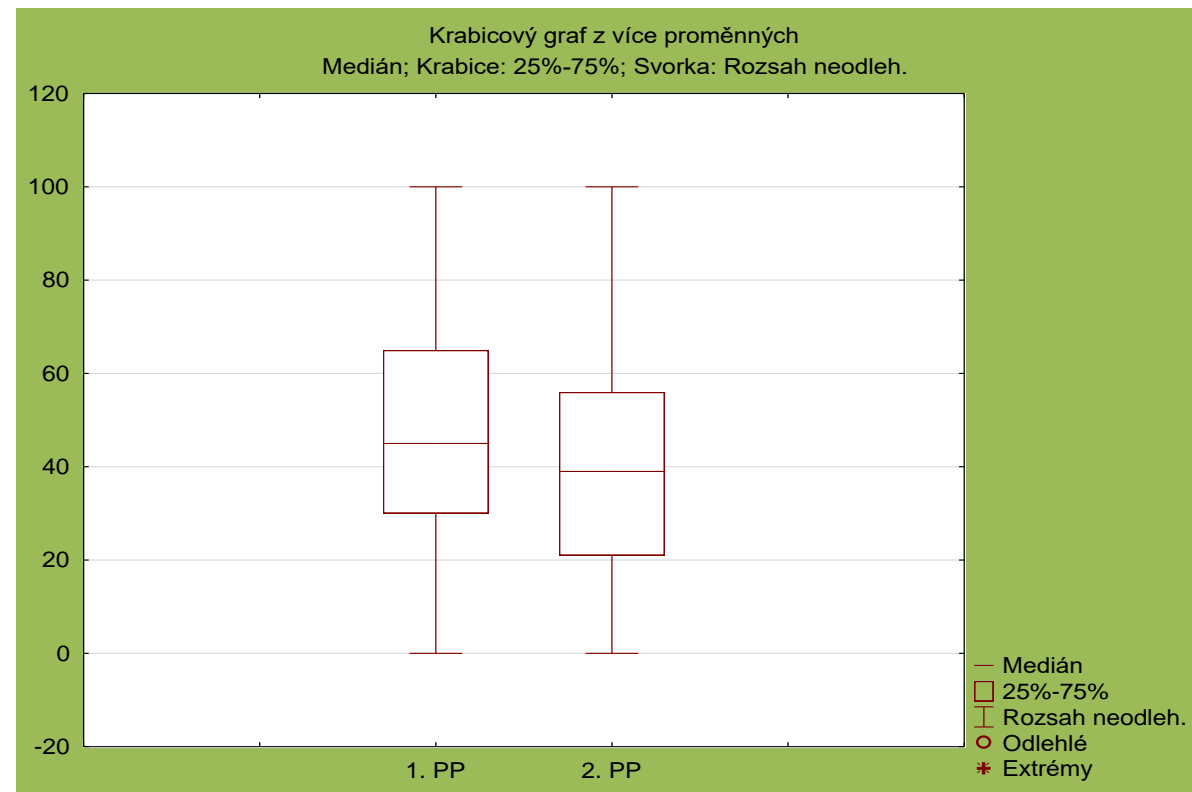
SELEKT CASES

Filtr případů

Váhy

Galerie grafů

Aktualizace: Auto



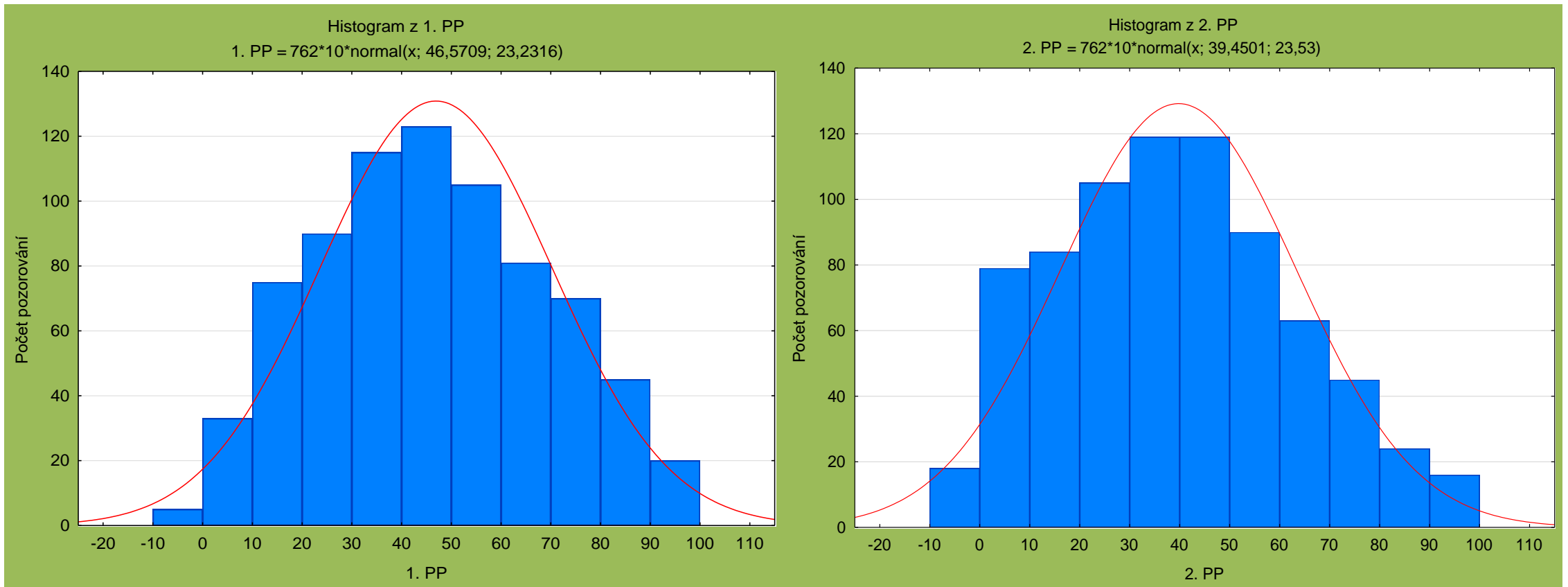
# Krabicový graf (box plot) II

- Krabicové grafy potvrzují naše předchozí úsudky na základě spočítaných charakteristik – v prvním testu dosahovali studenti o trochu více bodů, variabilita je v obou testech srovnatelná.
- Došlo také k malé změně rozdělení počtu bodů (viz posun mediánu uvnitř krabice u druhého testu).
- Vzhledem k tomu, že počet bodů je shora omezen 100 a zespoda 0, nejsou v datech žádná odlehlá pozorování.



# Histogram I

- Grafy → Histogram
- nastavení Proměnné 1. PP → OK → OK, stejně pro 2. PP



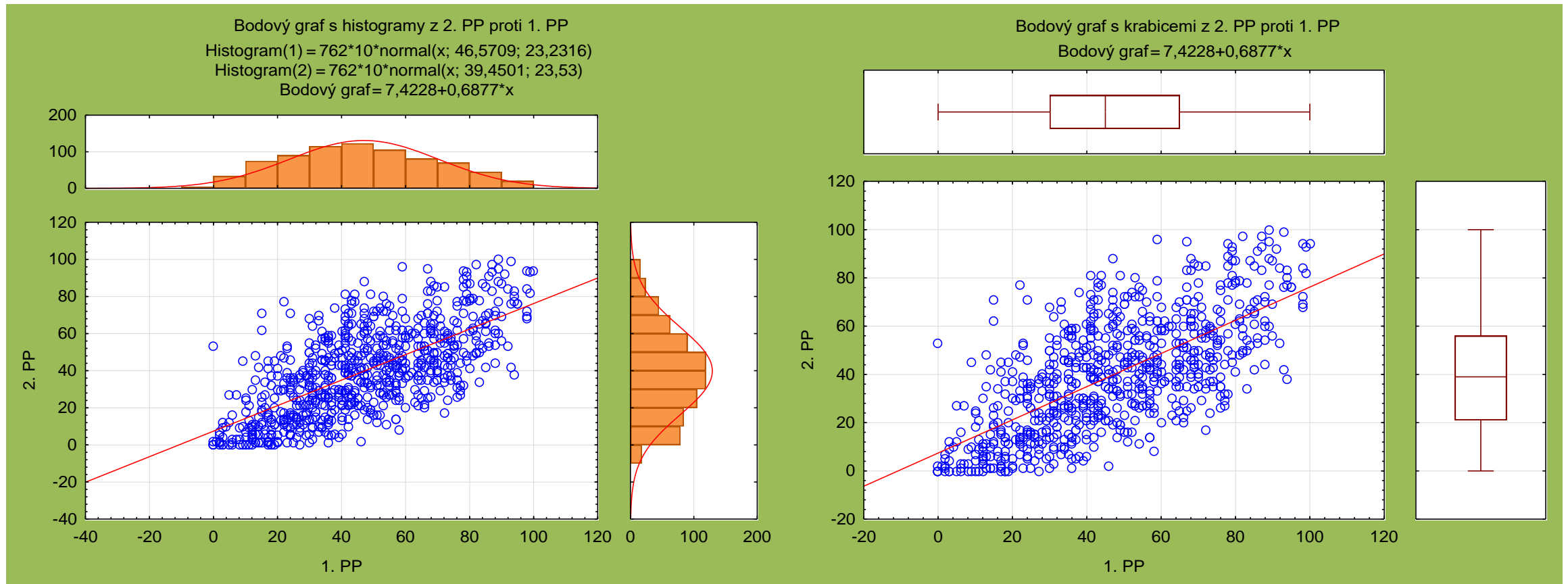
# Histogram II

- Rozdělení počtu bodů v prvním testu je dosti symetrické kolem středu (lze pozorovat jen velmi mírné pozitivní zešikmení) a neliší se příliš od Gaussovy křivky (v obrázku vyznačena červeně).
- Rozdělení počtu bodů v druhém testu je více pozitivně zešikmené. Nalevo od střední hodnoty jsou body rozděleny mnohem rovnoměrněji než napravo, kde četnosti klesají podobně jako Gaussova křivka.
- Oproti prvnímu testu můžeme pozorovat u druhého testu výrazné navýšení počtu studentů s velmi malým počtem bodů (0–10), a to je také jeden z hlavních důvodů poklesu průměrného počtu bodů.

# DALŠÍ GRAFICKÉ --- VÝSTUPY

# Bodový graf s histogramy

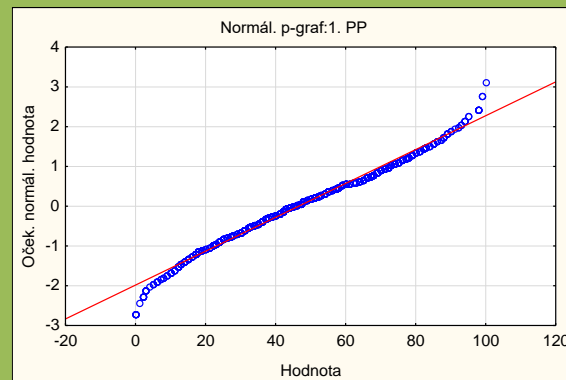
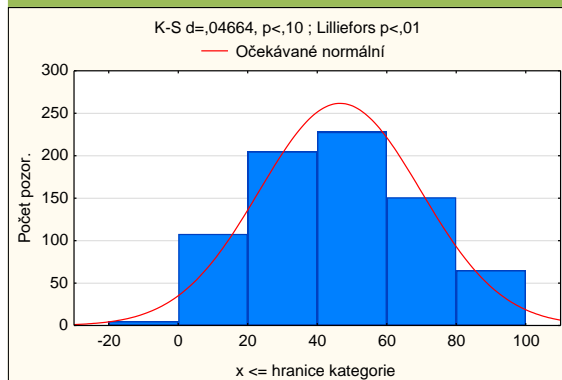
- Grafy → Grafy vstupních dat → Bodový graf dle → s histogramy či s krabicovými grafy



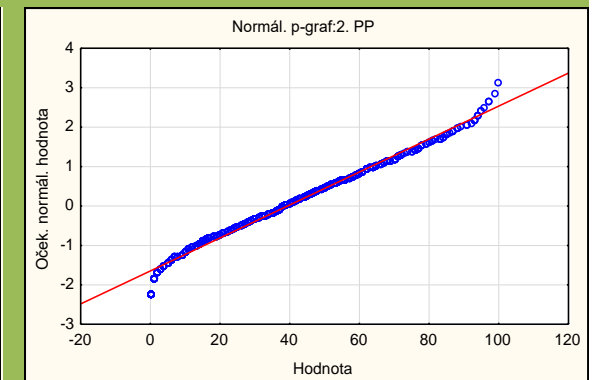
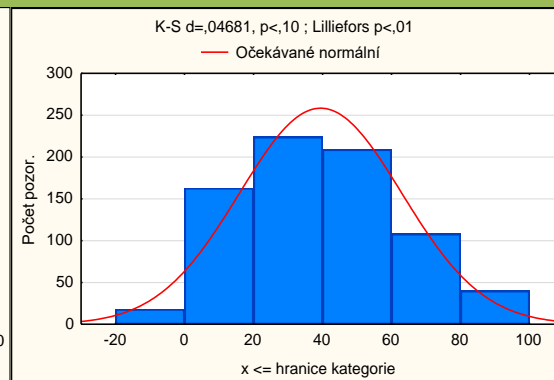
# Souhrnné výsledky

- **Statistiky** → Základní statistiky/tabulky → Základní statistiky/tabulky → zvolíme všechny proměnné → **Grafy 1**

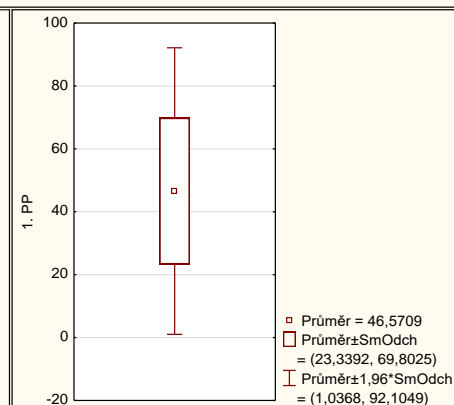
Souhrn: 1. PP



Souhrn: 2. PP



Souhrnné statistiky: 1. PP  
N platných=762,000000  
Průměr= 46,570866  
Minimum= 0,000000  
Maximum=100,000000  
Sm.odch.= 23,231641



Souhrnné statistiky:2. PP  
N platných=762,000000  
Průměr= 39,450131  
Minimum= 0,000000  
Maximum=100,000000  
Sm.odch.= 23,530018

