

## Cvičení 3: Základní pojmy matematické statistiky

### Úkol 1.: Průzkum chování výběrového průměru a výběrového rozptylu

1. Vytvořte nový datový soubor o 103 proměnných a 100 případech. Pomocí programu gener.svb, který si stáhnete z Učebních materiálů, se naplní prvních 100 proměnných 100 realizacemi náh. veličin  $X_i \sim R_s(0,1)$ ,  $i=1, \dots, 100$ , do proměnné v101 se uloží pořadová čísla 1 až 100, do proměnné v102 (resp. v103) se uloží průměry (resp. rozptyly) proměnných v1 až v100.

Option Base 1

Sub Main

Dim s As Spreadsheet

Set s = ActiveSpreadsheet

For i = 1 To 100

s.Variable(i).FillRandomValues

'do promennych v1 az v100 se ulozi nahodna cisla z intervalu(0,1)

Next i

s.VariableLongName(101) = "=v0"

'do promenne v101 se ulozi poradova cisla 1 az 100

s.VariableLongName(102) = "=mean(v1:v100)"

'do promenne v102 se ulozi prumery promennych v1 az v100

s.VariableLongName(103) = "=stdev(v1:v100)^2"

'do do promenne v103 se ulozi rozptyly promennych v1 az v100

s.Recalculate

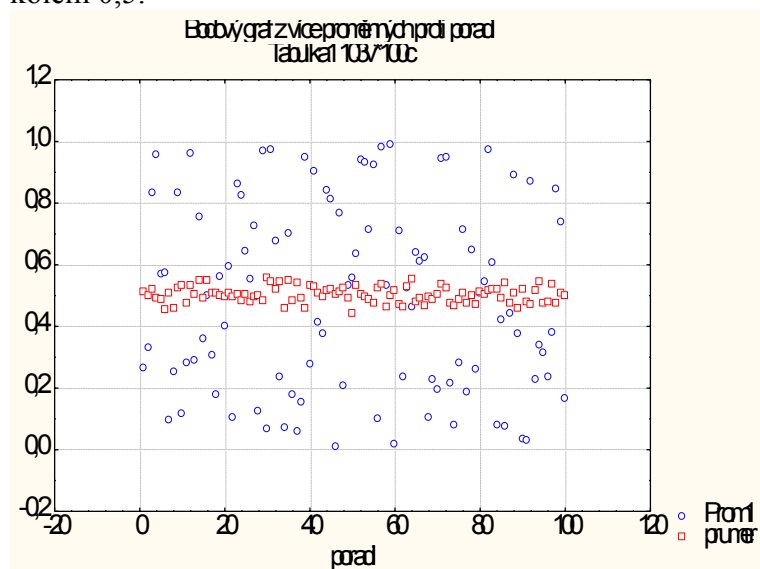
End Sub

(Makro se spouští pomocí modré šipky na panelu nástrojů.)

Proměnnou v101 přejmenujte na PORADI, v102 na PRUMER a v103 na ROZPTYL. Vzniklý datový soubor uložte pod názvem uniform.sta.

2. Graficky znázorněte hodnoty některé z proměnných v1, ..., v100 (např. v1) a hodnoty proměnné PRUMER.

Návod: Grafy – Bodové grafy – Typ grafu Vícenásobný – vypneme Lineární proložení – Proměnné X PORADI, Y v1, PRUMER, OK, OK. Vidíme, že hodnoty proměnné v1 se nacházejí mezi 0 a 1, zatímco hodnoty proměnné PRUMER se koncentrují v úzkém pásu kolem 0,5.

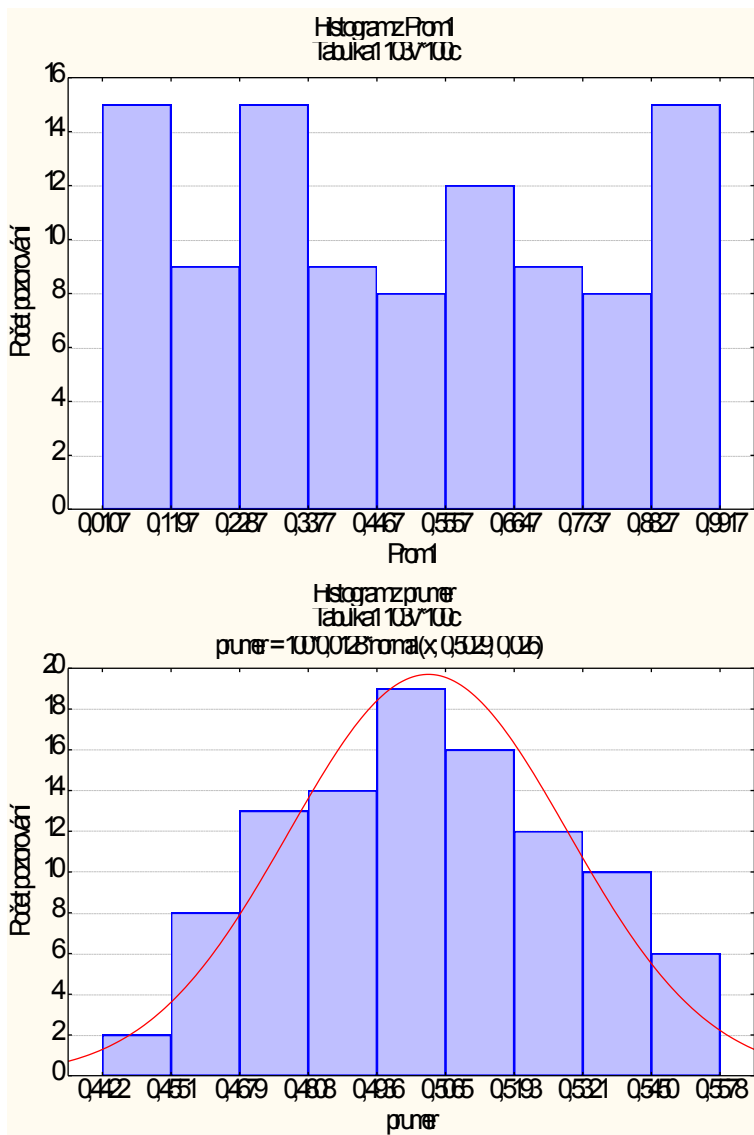


3. Vypočítejte průměr a rozptyl např. proměnné v1 a proměnné PRUMER. Průměr proměnné v1 by měl být blízký 0,5, rozptyl  $1/12 = 0,083$ . Průměr proměnné PRUMER by se měl blížit 0,5, zatímco rozptyl by měl být 100 x menší než  $1/12$ , tj. 0,00083. Dále vypočítejte průměr proměnné ROZPTYL. Měl by se blížit  $1/12 = 0,083$ .

Popis statistiky (unitom)		
Proměnná	Průměr	Rozptyl
Prum1	0,5366	0,0785
PRUMER	0,5029	0,0008

Popis statistiky (unitom)	
Proměnná	Průměr
ROZPTYL	0,08314

4. Nakreslete histogram pro proměnnou v1 a pro proměnnou PRUMER. První histogram se blíží úsečce, druhý Gaussově křivce.

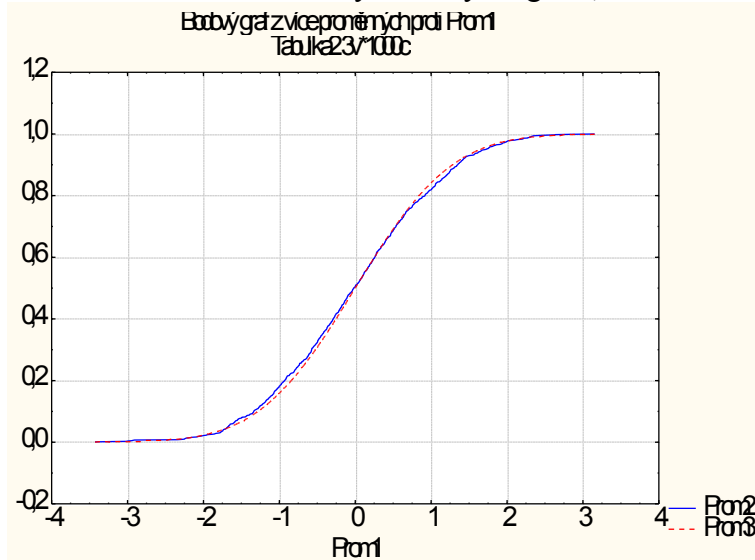


5. Celý postup zopakujte pro exponenciální rozložení s parametrem  $\lambda=2$ . V programu gener.stb napište místo s.Variable(i).FillRandomValues s.VariableLongName(i) = "=Vexpon(rnd(1);2) "

Připomeneme si, že průměr proměnné  $v_1$  i průměr proměnné PRUMER by se měl blížit  $1/2$ , rozptyl proměnné  $v_1$  by měl být blízký  $1/4$  a rozptyl proměnné PRUMER by měl být  $100$  x menší, tj.  $0,0025$ . Průměr proměnné ROZPTYL by se neměl příliš lišit od  $1/4$ .

## Úkol 2.: Ilustrace nestrannosti výběrové distribuční funkce

1. Vytvořte nový datový soubor o třech proměnných a 1000 případech.
2. Do proměnné  $v_1$  uložte 1000 realizací náhodné veličiny s rozložením  $N(0,1)$  tak, že v Dlouhém jménu použijte příkaz `=vnormal(rnd(1);0;1)`
3. Hodnoty proměnné  $v_1$  seřadíte podle velikosti: Data - Setřídít.
4. Proměnnou  $v_2$  transformujte tak, že v Dlouhém jménu použijte příkaz `=v0/1000`.
5. Do proměnné  $v_3$  uložte hodnoty distribuční funkce rozložení  $N(0,1)$ . Do Dlouhého jména napište příkaz `=INormal(v1;0;1)`
6. Nakreslete dvourozměrný tečkový diagram, kde na osu x vyneste  $v_1$  a na osu y  $v_2$  a  $v_3$ .



Vidíme, že průběh výběrové distribuční funkce  $F_{1000}(x)$  (modrá čára) je velmi podobný průběhu distribuční funkce  $\Phi(x)$  (červená čára).

7. Postup zopakujte pro rozsah výběru  $n = 100$ . Uvidíte, průběh výběrové distribuční funkce  $F_{100}(x)$  se od průběhu distribuční funkce  $\Phi(x)$  liší výrazněji.

## Úkol 3.: Sledování vlivu rozsahu výběru na šířku intervalu spolehlivosti (při $\alpha=0,05$ )

Pro hypotetické náhodné výběry rozsahu  $n$  ( $n = 5, 7, 9, \dots, 85$ ) z rozložení  $N(0,1)$ , jejichž výběrové průměry se vždy realizovaly hodnotou 0, vypočtete dolní a horní meze 95% intervalů spolehlivosti pro  $\mu$  a graficky znázorníte závislost těchto mezí na rozsahu  $n$ .

Upozornění: Meze  $100(1-\alpha)\%$  empirického intervalu spolehlivosti pro střední hodnotu při

známém rozptylu se počítají podle vzorců:  $d = m - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}$ ,  $h = m + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}$

**Návod:** Z Učebních materiálů stáhněte program intsp1.svb a otevřete ho v programovacím okně.

Option Base 1

Dim s As Spreadsheet

Sub Main

    alfa = 0.05

    'pevně zvolené riziko

    m = 0

    'pevně zvolený průměr

    sigma = 1

    'pevně zvolená směrodatná odchylka

    n = 3

    'počáteční rozsah výběru

    Set s = ActiveSpreadsheet

    For I = 1 To 41

        s.Cells(I, 2) = m - VNormal(1 - alfa / 2, 0, 1) / Sqrt(n + 2 \* I)

        'dolní mez intervalu spolehlivosti

        s.Cells(I, 3) = m + VNormal(1 - alfa / 2, 0, 1) / Sqrt(n + 2 \* I)

        'horní mez intervalu spolehlivosti

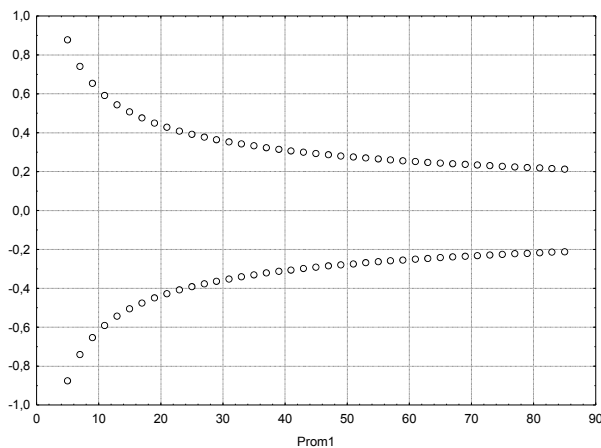
        s.Cells(I, 1) = n + 2 \* I

        'zvětšení rozsahu výběru o 2

    Next I

End Sub

Vytvořte nový datový soubor o 3 proměnných a 41 případech. Po spuštění programu intsp1 se do proměnné v1 uloží rozsahy výběrů 5, 7, ..., 85, do v2 (resp. v3) dolní (resp. horní) meze 95% intervalů spolehlivosti pro  $\mu$ . Vytvoření grafu: Grafy – Bodové grafy – Typ grafu Vícenásobný – vypneme Lineární proložení – Proměnné X v1, Y v2, v3 OK, OK.



Vidíme, že šířka intervalu spolehlivosti klesá se zvětšujícím se rozsahem náhodného výběru, zprvu rychle a pak stále pomaleji.

#### Úkol 4.: Sledování vlivu rizika na šířku intervalu spolehlivosti (při konstantním rozsahu výběru)

Pro hypotetický náhodný výběr rozsahu  $n=25$  z rozložení  $N(0,1)$ , jehož výběrový průměr se realizoval hodnotou 0, vypočítejte dolní a horní meze  $100(1-\alpha)\%$  intervalů spolehlivosti ( $\alpha=0,20, 0,19, \dots, 0,01$ ) pro  $\mu$  a graficky znázorněte závislost těchto mezí na riziku  $\alpha$ .

**Návod:** Z Učebních materiálů stáhněte program intsp2.svb a otevřete ho v programovacím okně.

Option Base 1

Dim s As Spreadsheet

Sub Main

alfa = 0.21

'počáteční hodnota rizika

m = 0

'pevně zvolený průměr

sigma = 1

'pevně zvolená směrodatná odchylka

n = 25

'pevně zvolený rozsah výběru

Set s = ActiveSpreadsheet

For I = 1 To 20

s.Cells(I, 2) = m - VNormal(1 - (alfa - I / 100) / 2, 0, 1) / Sqrt(n)

'dolní mez intervalu spolehlivosti

s.Cells(I, 3) = m + VNormal(1 - (alfa - I / 100) / 2, 0, 1) / Sqrt(n)

'horní mez intervalu spolehlivosti

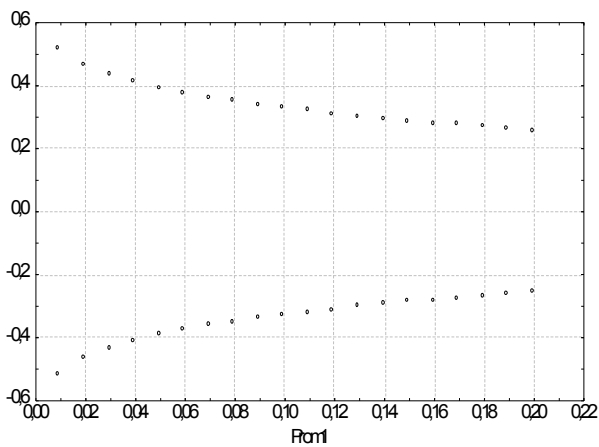
s.Cells(I, 1) = alfa - I / 100

'zmenšení rizika o 1/100

Next I

End Sub

Vytvořte nový datový soubor o 3 proměnných a 20 případech. Po spuštění programu intsp2 se do proměnné v1 uloží rizika 0,20, 0,19, ..., 0,01, do v2 (resp. v3) dolní (resp. horní) meze 100(1- $\alpha$ )% intervalů spolehlivosti pro  $\mu$ . Vytvoření grafu: stejným způsobem jako v předešlém případě.



Vidíme, že šířka intervalu spolehlivosti s rostoucím rizikem klesá.