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 náhodných veličin $X_i \sim Rs(0,1)$, i = 1, …, 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.
Zdrojový text programu gener.svb:

```
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
```

(Program se spouští pomocí modré šipky na panelu nástrojů.)
Proměnnou v101 přejmenujte na PORADI, v102 na PRUMER a v103 na ROZPTYL.

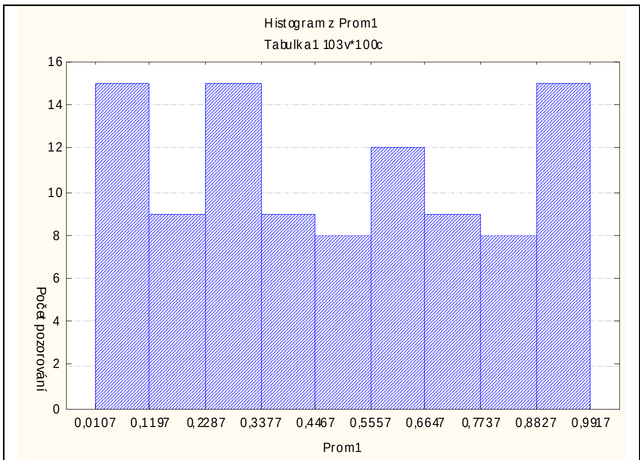
2. Graficky znázorněte hodnoty některé z proměnných v1, …, v100 (např. v1) a hodnoty proměnné PRUMER.
Návod: Grafy – Bodové grafy – Typ grafu Vícenásobný – vypneme Lineární proložení – Proměnné X PORADI, Y v1, PRUMER, OK, OK. Vidíme, že hodnoty proměnné v1 se nacházejí mezi 0 a 1, zatímco hodnoty proměnné PRUMER se koncentrují v úzkém pásu kolem 0,5. Znamená to, že průměr funguje jako těžiště dat - eliminuje příliš velké i příliš malé hodnoty.

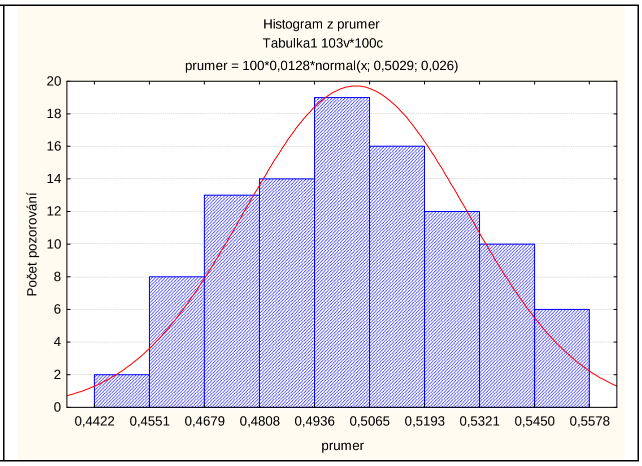
3. Vypočtěte průměr a rozptyl např. proměnné v1 a proměnné PRUMER. Průměr proměnné v1 by měl být blízký 0,5, rozptyl 1/12 = 0,083. Průměr proměnné PRUMER by se měl blížit 0,5, zatímco rozptyl by měl být 100 x menší než 1/12, tj. 0,00083. Dále vypočtěte průměr proměnné ROZPTYL. Měl by se blížit 1/12 =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

	Popisné statistiky (uniform)	
Proměnná	Průměr	Rozptyl
Prom1	0,536605	0,078676
PRUMER	0,503984	0,000783

	Popisné statistiky (uniform)
Proměnná	Průměr
ROZPTYL	0,083143

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





Úkol 2.: Ilustrace nestrannosti výběrové distribuční funkce

1. Vytvořte nový datový soubor o třech proměnných a 1000 případech.

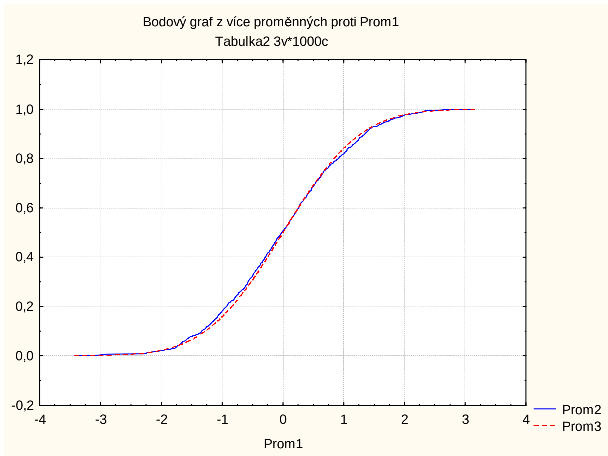
2. Do proměnné v1 uložte 1000 realizací náhodné veličiny s rozložením N(0,1) tak, že v Dlouhém jménu použijte příkaz =vnormal(rnd(1);0;1)

3. Hodnoty proměnné v1 setřiďte podle velikosti: Data - Setřídit.

4. Proměnnou v2 transformujte tak, že v Dlouhém jménu použijte příkaz =v0/1000. Tím získáme hodnoty výběrové distribuční funkce.

5. Do proměnné v3 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 v1 a na osu y v2 a v3.



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, že průběh výběrové distribuční funkce $F_{100}(x)$ se od průběhu distribuční funkce $\Phi(x)$ liší výrazněji.

Úkol 3.: Vlastnosti výběrového průměru z normálního rozložení
Předpokládejme, že velký ročník na vysoké škole má výsledky ze statistiky normálně rozloženy kolem střední hodnoty 72 bodů se směrodatnou odchylkou 9 bodů. Najděte pravděpodobnost, že průměr výsledků náhodného výběru 10 studentů bude větší než 80 bodů.
Návod:
$X_1, ..., X_{10}$ je náhodný výběr z N(72, 81). Počítáme P(M > 80), přičemž výběrový průměr M má normální rozložení se střední hodnotou E(M) = μ = 72 a rozptylem D(M) $= \frac{\sigma^2}{n} = \frac{81}{10} =$ 8,1. Tedy P(M > 80) = 1 - P(M $\leq$ 80) = 1 – Φ(80), kde Φ(80) je hodnota distribuční funkce rozložení N(72; 8,1) v bodě 80.
Výpočet pomocí systému STATISTICA: Vytvoříme datový soubor o jedné proměnné a o jednom případu. Do Dlouhého jména této proměnné napíšeme =1 – INormal(80;72;sqrt(8,1)). Zjistíme, že 1 - Φ(80) = 0,00247005. Funkce INormal(x;μ;σ) počítá hodnotu distribuční funkce rozložení N(μ,σ^2) v bodě x.

Úkol 4.: Výpočet mezí intervalu spolehlivosti pro střední hodnotu normálního rozložení při známém rozptylu
Rychlost letadla byla určována v pěti zkouškách. Z jejich výsledků byl vypočten průměr m = 870,3 m/s. Z dřívějších měření je známo, že rychlost letadla se řídí normálním rozložením se směrodatnou odchylkou 2,1 m/s. Najděte 95% empirický interval spolehlivosti pro neznámou střední hodnotu rychlosti μ.
Upozornění: Meze 100(1-α)% 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: Vytvoříme nový datový soubor o dvou proměnných a jednom případu. Proměnnou v1 pojmenujeme DM, v2 HM. Do Dlouhého jména proměnné DM napíšeme
=870.3-2.1*VNormal(0.975;0;1)/sqrt(5)
a do Dlouhého jména proměnné HM napíšeme
=870.3+2.1*VNormal(0.975;0;1)/sqrt(5).
Dostaneme výsledek $\mu \in (868,5;872,1)$ s pravděpodobností aspoň 0,95.