Téma 9: Centrální limitní věta

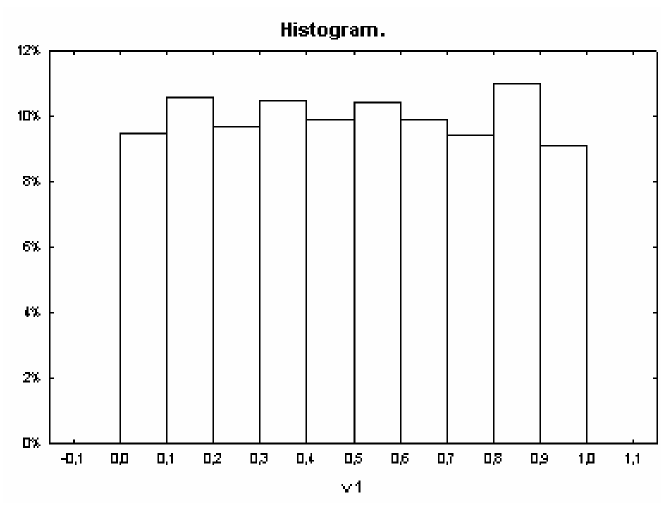
Ilustrace centrální limitní věty

Vygenerujeme 12 x 1000 realizací náhodných veličin $X_1, ..., X_{12}$, $X_i \sim Rs(0,1)$, i =1, ..., 12. Podle centrální limitní věty má náhodná veličina $X = X_1 + ... + X_{12} - 6$ přibližně rozložení N(0,1).
Návod: Vytvoříme nový datový soubor o 13 proměnných a 1000 případech. Otevřeme programovací okno STATISTICA VISUAL BASIC (File – New – Macro (SVB) Program – Name clv – OK) a do okna napíšeme příkazy:

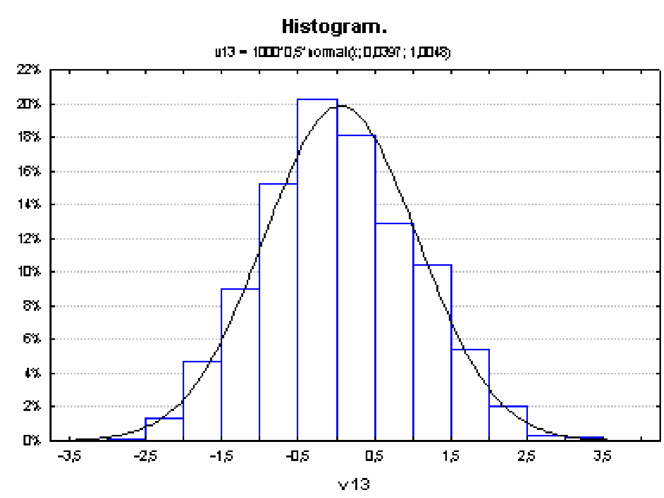
```
Dim s As Spreadsheet
Set s = ActiveSpreadsheet
For i = 1 To 12
  s.Variable(i).FillRandomValues
  ' do proměnných v1 až v12 se uloží náhodná čísla
  ' z intervalu(0,1)
Next i
s.VariableLongName(13) = "=Sum(v1:v12)-6"
' do proměnné v13 se uloží součet proměnných v1 až v12
' zmenšený o 6
s.Recalculate
```

Znázorníme histogramy proměnných v1 a v13 a porovnáme jejich vzhled s tvarem hustot rozložení Rs(0,1), N(0,1).
Dále spočteme průměry a rozptyly proměnných v1 a v13 a porovnáme je s teoretickou střední hodnotou a rozptylem náhodné veličiny s rozložením Rs(0,1) (E(X)=0,5, D(X)=1/12=0,833) a náhodné veličiny s rozložením N(0,1) (E(X)=0, D(X)=1).

Řešení:



Jedná se o 1000 náhodných čísel vygenerovaných z intervalu (0,1).
Jejich aritmetický průměr je m = 0,497491 a rozptyl s^2 = 0,082374.
Střední hodnota Rs(0,1) je E(X) = 0,5 a rozptyl D(X) = 1/12 = 0,08333.
Sloupky v histogramu by měly lehce kolísat kolem hodnoty 100, protože při rovnoměrném spojitém rozložení na intervalu (0, 1) by absolutní četnost každého z intervalů (0, 1/10], (1/10, 2/10], …, (9/10, 1) měla být 100.



Variable	Descriptive Statistics	
	Mean	Variance
v1	0,497471	0,082374
v13	0,039656	1,009721

Jedná se o náhodnou veličinu v13 = v1+v2+...+v12 – 6, která podle centrální limitní věty má přibližně rozložení N(0,1). (Přesněji řečeno, posloupnost standardizovaných součtů konverguje v distribuci ke standardizované normální náhodné veličině.)
Aritmetický průměr v13 vyšel m = 0,039656, rozptyl s^2 = 1,009721.
Střední hodnota X ~ N(0,1) je E(X) = 0, rozptyl D(X) = 1.

(Při každém provedení tohoto úkolu dostaneme trochu jiné výsledky, protože postup je založen na generování náhodných čísel.)

Aplikace Moivreovy - Laplaceovy integrální věty

Pomocí STATISTIKY spočteme př. 11.2. ze skript Teorie pravděpodobnosti a matematická statistika: Pravděpodobnost úspěchu při jednom pokusu je 0,3. S jakou pravděpodobností lze tvrdit, že počet úspěchů ve 100 pokusech bude v mezích od 20 do 40?

Y_{100} – počet úspěchů v posloupnosti n = 100 opakovaných nezávislých pokusů, pravděpodobnost úspěchu $\vartheta = 0,3$, $Y_{100} \sim$ Bi(100, 0,3), $E(Y_{100}) = n\vartheta = 30$, $D(Y_{100}) = n\vartheta(1-\vartheta) = 21$.

Aproximativní výpočet:

$$P(20 \le Y_{100} \le 40) = P\left(\frac{19-30}{\sqrt{21}} < \frac{Y_{100}-30}{\sqrt{21}} \le \frac{40-30}{\sqrt{21}}\right) \approx \Phi\left(\frac{10}{\sqrt{21}}\right) - \Phi\left(-\frac{11}{\sqrt{21}}\right) = 0,9773,$$

kde $\Phi(x)$ je distribuční funkce rozložení N(0,1).

Postup ve STATISTICE:
File – New – Number of variables 2, Number of cases 1 – OK.
Nastavíme se kurzorem na 1. sloupec.
Long Name =INormal(10/sqrt(21);0;1)- INormal(-11/sqrt(21);0;1) OK. (Funkce INormal(x;mu;sigma) poskytuje hodnotu distribuční funkce v bodě x normálního rozložení se střední hodnotou mu a směrodatnou odchylkou sigma.)

Přesný výpočet:

$$P(20 \le Y_{100} \le 40) = P(19 < Y_{100} \le 40) = \Phi(40) - \Phi(19) = 0{,}978614,$$

kde $\Phi(x)$ je distribuční funkce rozložení Bi(100, 0,3).

Postup ve STATISTICE:

Nastavíme se kurzorem na 2. sloupec.

Long Name =IBinom(40;0,3;100)- IBinom(19;0.3;100). (Funkce IBinom(x;p;n) poskytuje hodnotu distribuční funkce v bodě x binomického rozložení s parametry p a n.)

Podle tohoto návodu vyřešte příklady 11.3., 11.5., 11.6.

Př. 11.3.: Pravděpodobnost, že zakoupený elektrospotřebič bude vyžadovat opravu během záruční doby, je rovna 0,2. Jaká je pravděpodobnost, že během záruční doby bude nutno ze 400 prodaných spotřebičů opravit více než 96?

n = 400, ϑ = 0,2, úspěch je nutnost opravy v záruční době

$n\vartheta = 80$, $n\vartheta(1-\vartheta) = 64$

aproximativní výpočet: $P(Y_{400} > 96) \approx$ 1 - INormal(16/8;0;1) = 0,022750

přesný výpočet: $P(Y_{400} > 96) =$ 1 – IBinom(96;0,2;400) = 0,024640

Př. 11.5.: Pravděpodobnost narození chlapce je 0,515. Jaká je pravděpodobnost, že mezi 10 000 novorozenci bude

a) stejně nebo více děvčat než chlapců
b) chlapců od 5000 do 5300?

n = 10000, ϑ = 0,515, úspěch je narození chlapce

$n\vartheta = 5150$, $n\vartheta(1-\vartheta) = 2497{,}75$

Úkol (a)

aproximativní výpočet: $P(Y_{10000} \le 5000) \approx$ INormal(-150/sqrt(2497,75);0;1) = 0,001344

přesný výpočet: $P(Y_{10000} \le 5000) =$ IBinom(5000;0,515;10000) = 0,001347

Úkol (b)

aproximativní výpočet: $P(4999 < Y_{10000} \le 5300) \approx$ INormal(150/sqrt(2497,75);0;1) - INormal(-151/sqrt(2497,75);0;1) = 0,997399

přesný výpočet: $P(4999 < Y_{10000} \le 5300) =$ IBinom(5300;0;1) – IBinom(4999;0;1) = 0,997400

Př. 11.6.: Pravděpodobnost, že určitý typ výrobku má výrobní vadu, je 0,05. Jaká je pravděpodobnost, že ze série 1000 výrobků bude mít výrobní vadu nejvýše 70?

n = 1000, ϑ = 0,05, úspěch je zhotovení vadného výrobku

$n\vartheta = 50$, $n\vartheta(1-\vartheta) = 47{,}5$

aproximativní výpočet: $P(Y_{1000} \le 70) \approx$ INormal(20/sqrt(47,5);0;1) = 0,998145

přesný výpočet: $P(Y_{1000} \le 70) =$ IBinom(70;0,05;1000) = 0,997670