

Pomůcka pro cvičení: 3. semestr Bc studia

Statistika

Základní pojmy

balíček: Statistics

Pro veškeré výpočty je třeba načíst balíček **Statistic**. Při řešení můžeme použít proceduru **infolevel[Statistics]:=1**, která nám poskytne podrobný výpis informací vztahující se k danému výpočtu.

Dále budou uvedeny a na příkladech ukázány jen některé základní pojmy z přednášek. Balíček **Statistic** je výrazně bohatší, případné zájemce o další zpracování statistického materiálu odkazují na Help.

Příklad.

```
> with(Statistics):  
> infolevel[Statistics]:=1;  
infolevelStatistics := 1  
  
>  
data:=[6.5,6.2,5.5,5.25,4.8,4.75,4.2,3.5,1.5,1.4,0.75,0.575,0.5,0  
.46,0.35,0.315,0.29,0.1425,0.1375,0.135,0.125,0.1115,0.1115,0.109  
,0.109,0.109];  
data := [6.5, 6.2, 5.5, 5.25, 4.8, 4.75, 4.2, 3.5, 1.5, 1.4, 0.75, 0.575, 0.5,  
0.46, 0.35, 0.315, 0.29, 0.1425, 0.1375, 0.135, 0.125, 0.1115,  
0.1115, 0.109, 0.109, 0.109]
```

Příkaz **sort** setřídí zadaná data do neklesající posloupnosti.

```
> sort(data);  
[0.109, 0.109, 0.109, 0.1115, 0.1115, 0.125, 0.135, 0.1375, 0.1425,  
0.29, 0.315, 0.35, 0.46, 0.5, 0.575, 0.75, 1.4, 1.5, 3.5, 4.2, 4.75, 4.8,  
5.25, 5.5, 6.2, 6.5]
```

Příkaz **min** najde nejmenší x_{\min} a příkaz **max** největší x_{\max} hodnotu v zadaných datech.

```
> min(data);  
0.109  
  
> max(data);  
6.5
```

Variační rozpětí $R = x_{\max} - x_{\min}$ nalezneme pomocí příkazu **Range**.

```
> Range(data);  
6.391000000
```

Výběrový průměr $x_p = \frac{1}{n} \sum_{i=1}^n x_i$ nalezneme příkazem **Mean**.

```
> Mean(data);  
1.843461538
```

Rozptyl $s^2 = \frac{1}{(n-1)} \left[\sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2 \right]$ stanovíme příkazem **Variance**.

```
> Variance(data);
```

5.26568033846154

Směrodatná odchylka $s = \sqrt{s^2}$.

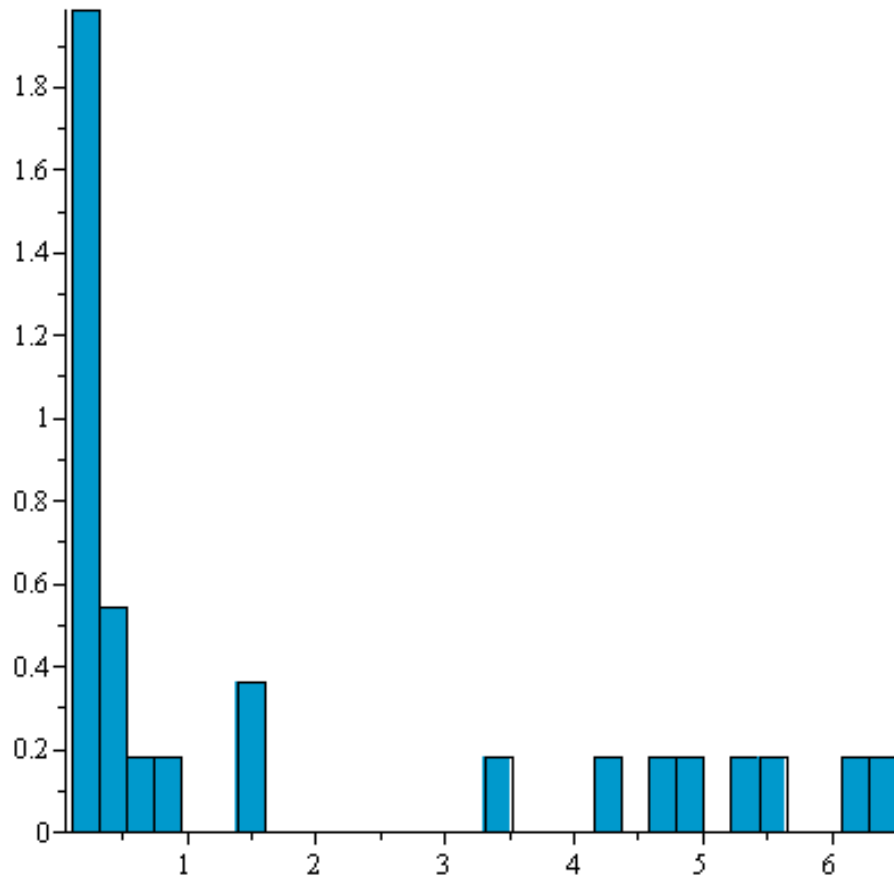
```
> StandardDeviation(data);
```

2.29470702671638

Pro nalezení histogramu použijeme příkaz **Histogram**.

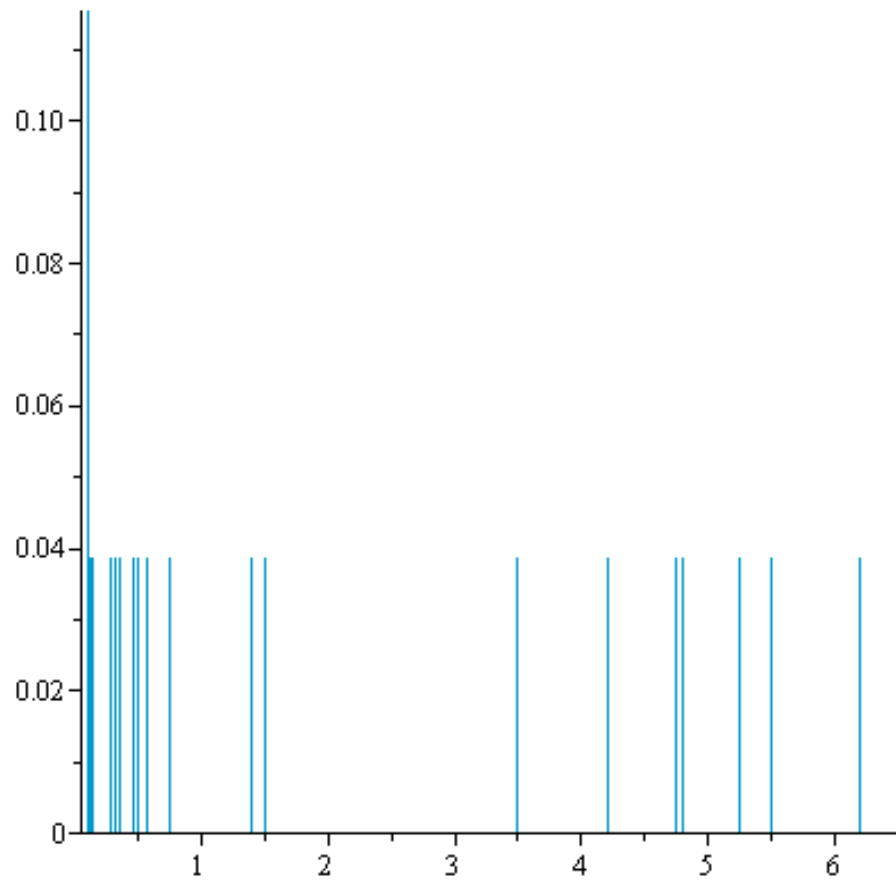
```
> Histogram(data);
```

```
Histogram Type: default  
Data Range:      .1090000000 .. 6.500000000  
Bin Width:       .2130333333  
Number of Bins:  30  
Frequency Scale: relative
```



```
> Histogram(data,discrete=true);
```

```
Histogram Type: discrete  
Data Range:      .1090000000 .. 6.500000000  
Number of Bins:  23  
Frequency Scale: relative
```



V případě, že chceme změnit počet intervalů v histogramu, je potřeba přidat parametr **bincount=n**, kde n je celé číslo.

```
> Histogram(data, bincount=5, frequencyscale = absolute);
```

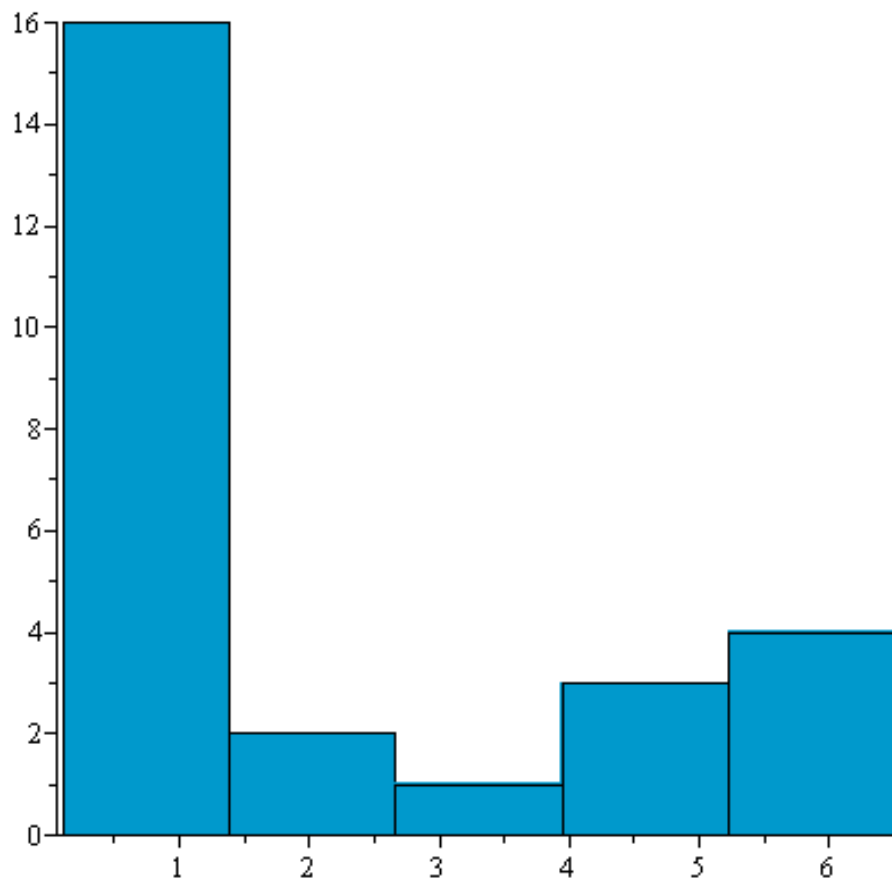
```
Histogram Type: default
```

```
Data Range: .1090000000 .. 6.500000000
```

```
Bin Width: 1.278200000
```

```
Number of Bins: 5
```

```
Frequency Scale: absolute
```



Rozdělení datového souboru do tříd. Počet tříd je dífoltně nastaven na 10.

FrequencyTableobsahuje 5 sloupců. V prvním je variační rozpětí, ve druhém the absolute frequency, ve třetím the percentage, ve čtvrtém the cumulative frequency a v pátém the cumulative percentage of the data.

> FrequencyTable (data) ;

```

0.1090000000 ..0.7481000000 15. 57.69230769 15. 57.69230769
0.7481000000 ..1.3872000000 1. 3.846153846 16. 61.53846154
1.3872000000 ..2.0263000000 2. 7.692307692 18. 69.23076923
2.0263000000 ..2.6654000000 0. 0. 18. 69.23076923
2.6654000000 ..3.3045000000 0. 0. 18. 69.23076923
3.3045000000 ..3.9436000000 1. 3.846153846 19. 73.07692308
3.9436000000 ..4.5827000000 1. 3.846153846 20. 76.92307692
4.5827000000 ..5.2218000000 2. 7.692307692 22. 84.61538462
5.2218000000 ..5.8609000000 2. 7.692307692 24. 92.30769231
5.8609000000 ..6.5000000000 2. 7.692307692 26. 100.00000000

```

Pokud chceme změnit počet tříd, je potřeba použít příkaz **bins=n**, kde n je celé číslo.

> FrequencyTable (data, bins=5) ;

0.1090000000	..1.387200000	16.	61.53846154	16.	61.53846154
1.387200000	..2.665400000	2.	7.692307692	18.	69.23076923
2.665400000	..3.943600000	1.	3.846153846	19.	73.07692308
3.943600000	..5.221800000	3.	11.53846154	22.	84.61538462
5.221800000	..6.500000000	4.	15.38461538	26.	100.0000000

Medián je 50%-ní kvantil. Je používán v těch případech, kdy náhodná veličina nemá definovanu střední hodnotu. Obecně není určen jednoznačně.

> **Median(data) ;**

0.4800000000

Šikmost.

> **Skewness(data) ;**

0.910065788271881

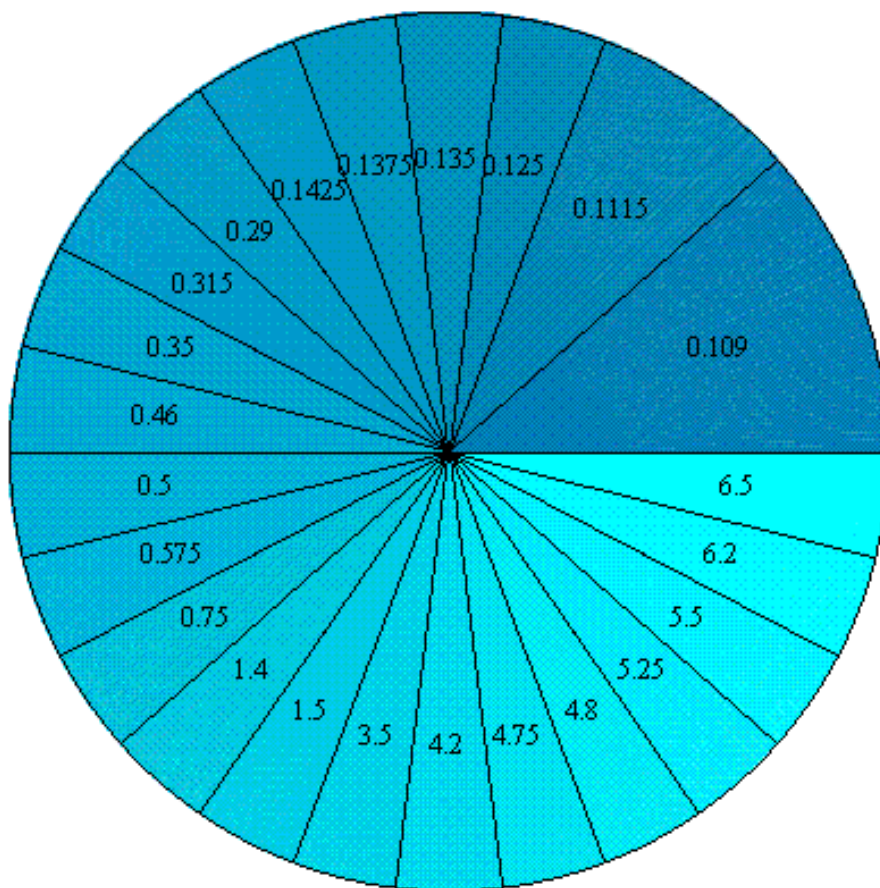
The Quantile function computes the quantile corresponding to the given probability p for the specified random variable or data set.

> **Quantile(data,1/3) ;**

0.1588888889

Data lze znázornit různými diagramy, jedním z nich je např. kruhový (koláčový) diagram.

> **PieChart(data) ;**



>

Generování náhodné proměnné

balíček: Statistics

Náhodnou proměnnou generujeme pomocí příkazu **RandomVariable(typ rozdělení)**, kde napíšeme o jaké rozdělení a s jakými parametry by se mělo jednat. Pokud chceme vidět vygenerované hodnoty, je potřeba použít příkazu **Sample(název náh. prom., počet vygenerovaných prvků)**.

Následně pak můžeme počítat veškeré charakteristiky, které potřebujeme.

Př. 1 Obchodní cestující prodává pračky. Na obchodní cesty jezdí se čtyřmi pračkami. Statisticky má zjištěno, že průměrně dva z devíti zákazníků, kterým pračku nabídne, si ji koupí. Chce odhadnout pravděpodobnosti pro počet prodaných praček a střední hodnotu tohoto počtu po čtyřech nabídkách.

Zřejmě jde o binomické rozdělení s parametry $n=4$ a $p=2/9$.

```
> with(Statistics):  
> R := RandomVariable(Binomial(4, 2/9));  
R := _R
```

Pokud bychom chtěli vidět jednu z možností prodeje obchodního cestujícího během 10 cest, je možné si je nechat vygenerovat.

```
> S:=Sample(R,10);  
S:= [ 2. 2. 0. 2. 1. 0. 0. 1. 2. 3. ]
```

Pokud chceme zjistit s jakou pravděpodobností prodá na svých cestách jednu pračku, využijeme příkazu **ProbabilityFunction**. Zde do závorky zadáme jméno náhodné proměnné, uvedeme v jakém čísle chceme pravděpodobnost určit a jako nepovinný parametr zadáme numeric, tj. výsledek bude zapsán desetinným číslem.

```
> f:=ProbabilityFunction(R, 1,numeric);  
f:= 0.4182289285
```

POZN: Pro zajímavost si můžete nechat vygenerovat několikrát úspěšný prodej obchodního cestujícího, zjistíte, že číslo 4, tj. prodej čtyř praček se v něm téměř nevyskytne. Určete si pravděpodobnost prodeje čtyř praček.

Střední hodnotu a rozptyl počtu prodaných praček určíme následovně:

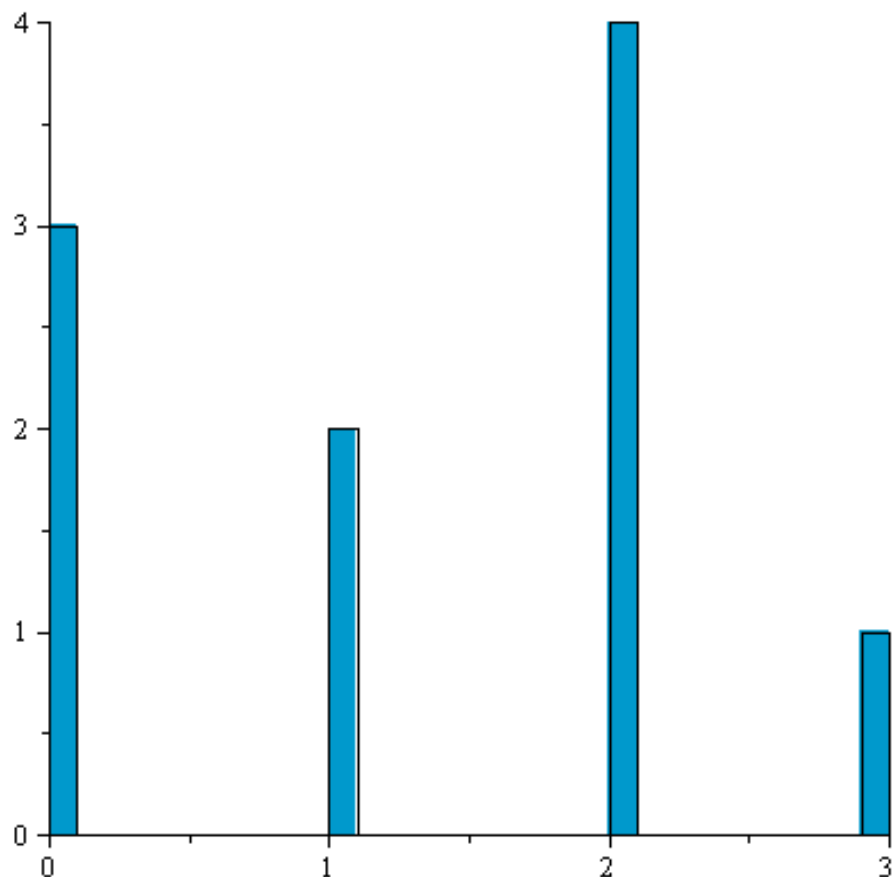
```
> Mean(R);  

$$\frac{8}{9}$$
  
> Variance(R);  

$$\frac{56}{81}$$

```

```
> Histogram(S);  
Histogram Type: default  
Data Range: 0. .. 3.  
Bin Width: .1000000000  
Number of Bins: 30  
Frequency Scale: relative
```



Př. 2 Stroj, který vyrábí součásty, je seřízen tak, aby střední hodnota jejich délek byla 42 mm. Přesnost je taková, že směrodatná odchylka délky součástek je 1,2 mm. Vygenerujte sérii deseti součástek a spočítejte s jakou pravděpodobností bude vyrobena součástka délky 44,3 mm.

> **N:=RandomVariable(Normal(42,1.2));**

N := _R0

> **S1:=Sample(N,10);**

*S1 := [42.2446264828453, 41.8813538111675, 41.9007622835496,
41.7982570462543, 42.1944255598415, 41.1888975483929,
41.4710428271418, 43.0752239823115, 42.5532275440560,
43.5335724627286]*

> **ProbabilityDensityFunction(Normal(a,b),t);**

$$\frac{1}{\sqrt{2\pi}b} e^{-\frac{1}{2} \frac{(t-a)^2}{b^2}}$$

> **ProbabilityDensityFunction(N,t);**

$$\frac{0.4166666667 \sqrt{2} e^{-0.3472222222 (t-42)^2}}{\sqrt{\pi}}$$

> **ProbabilityDensityFunction(N,44.3);**

0.05296808873

```
> ProbabilityDensityFunction(N,42,numeric);  
0.3324519002  
>
```