Oprava 8.4.2009

Statistika v Maple pro Y01PST

Základní příkazy

$abs(a)$	$\|a\|$;
$binomial(n,k)$	$\binom{n}{k}$
$n!$ nebo $factorial(n)$	$n!$, faktoriál
$exp(x)$	e^x
$ln(x)$	$\ln x$
$sqrt(x)$	$\sqrt{x}$
$surd(x,n)$	$\sqrt[n]{x}$
$sum(a(k), k = k1..k2)$	$\sum\limits_{k=k_1}^{k_2} a_k$
$limit(a(n), n = infinity)$	$\lim\limits_{n\to\infty} a_n$
$limit(f, x = a)$	$\lim\limits_{x\to a} f(x)$ funkce
$limit(f(x), x = a)$	$\lim\limits_{x\to a} f(x)$ výraz
$int(f, x)$	$\int f(x)\,\mathrm{d}x$ funkce
$int(f(x), x)$	$\int f(x)\,\mathrm{d}x$ výraz
$int(f, x = a..b)$	$\int_a^b f(x)\,\mathrm{d}x$ funkce
$int(f(x), x = a..b)$	$\int_a^b f(x)\,\mathrm{d}x$ výraz
$D(f)$	$f'(x)$ funkce
$diff(f(x), x)$	$f'(x)$ výraz

Zadání funkce

a) **jako funkce:** Maple rozumí symbolu $f(x)$, $f(1)$

$f := x-> vzorec \quad f(x);$

zadání po úsecích

$f := x-> piecewise(kde, vzorec, kde, vzorec, ...)$

b) **jako výraz:**

$f := vzorec$ pro $f(x)$; x není proměnná $f(1)$ dostaneme jako $x = 1 : f$; nebo $f(1) = subs(x = 1, f)$

zadání po úsecích

$f := piecewise(kde, vzorec, kde, vzorec, ...)$

Kreslení grafu

$plot(f, x = a..b)$ funkce $plot(f(x), x = a..b)$ výraz

Zadání dalších parametrů, příklad funkce (pro výraz $f(x)$)

$A := plot(f, a..b, discont = true, thickness = 1, colour = red) : plots[display](\{A\});$

$discont$ je parametr pro nespojité funkce

$thickness$ tlouška čáry (1,2,3)

$colour$ barva (red, black, blue, green, yellow)

Pro posloupnosti

$A := plot(a, 1..N, style = point, symbol = box, colour = red) : plots[display](\{A\});$

$symbol$ box, circle, diamond

Speciální funkce

$Beta(p, q)$ Eulerova funkce Beta $B(p, q)$

$Gamma(x)$ Eulerova funkce gamma $\Gamma(x)$

Výpočty hodnot rozdělení

hustota, resp. pravděpodobnostní funkce, distribuční funkce, kvantily

$with(stats):$

$statevalf[1,2](x)$

1. druh funkce:

a) spojité rozdělení:

cdf	distribučnífunkce
icdf	kvantil
pdf	hustota

b) diskrétní rozdělení:

dcdf	distribuční funkce
idcdf	kvantil
pf	pravděpodobnostní funkce

2. typ rozdělení

a) spojité rozdělení:

$beta[p,q]$	Beta rozdělení $B(p,q)$;
$cauchy[a,b]$	Cauchyovo rozdělení $C(a,b)$;
$chisquare[n]$	rozdělení $\chi^2(n)$;
$exponential[A,\delta]$	exponenciální rozdělení $Exp(A,\delta)$;
$fratio[m,n]$	Fischerovo rozdělení $F(m,n)$ -podíl χ^2;
$gamma[m,\delta]$	rozdělení gamma $\Gamma(m,\delta)$;
$laplaced[a,b]$	Laplaceovo rozdělení $L(a,b)$;
$normald[\mu,\sigma]$	normální rozdělení $N(\mu,\sigma^2)$;
$normald$	normované normální rozdělení $N(0,1)$;
$studentst[n]$	Studentovo $t-$rozdělení $t(n)$;
$uniform[a,b]$	rovnoměrné rozdělení v intervalu (a,b);
$uniform$	rovnomorné rozdělení v intervalu $(0,1)$;
$weibull[c,\delta]$	Weibullovo rozdělení $W(c,\delta)$.

b) diskrétní rozdělení:

$binomial[n,p]$	binomické rozdělení $Bi(n,p)$;
$discreteuniform[a,b]$	rovnoměrné diskrétní v intervalu (a,b);
$hypergeometric[N,M,n]$	hypergeometrické rozdělení s parametry N,M,n;
$poisson[\mu]$	Poissonovo rozdělení $Po(\mu)$.

Generování hodnot náhodné veličiny s požadovaným rozdělením

$with(stats)$:

$random$[druh rozdělení](počet hodnot) :

druh rozdělení je možné volit:

$beta[p,q]$	Beta rozdělení $B(p,q)$;
$chisquare[n]$	rozdělení $\chi^2(n)$;
$fratio[m,n]$	Fischerovo rozdělení $F(m,n)$ -podíl χ^2;
$gamma[m,\delta]$	rozdělení gamma $\Gamma(m,\delta)$;
$normald[\mu,\sigma]$	normální rozdělení $N(\mu,\sigma^2)$;
$normald$	normované normální rozdělení $N(0,1)$;
$studentst[n]$	Studentovo $t-$rozdělení $t(n)$;

Pro ostatní běžná rozdělení se hodnoty náhodné veličiny generují z rovnoměrného rozdělení pomocí vzorce pro kvantil.

Popisná statistika

Zadání dat

data:=[item, item,…, item]

item = číslo;

item = $Weight(m,n)$ - hodnotu m obsahují data $n-$krát;

item = $Weight(m_1..m_2)$ - je hodnota $\frac{1}{2}(m_1+m_2)$;

item = $Weight(m_1..m_2,n)$ - je n hodnot $\frac{1}{2}(m_1+m_2)$; (nepoužívat)

item = $missing$ - údaj chybí

Lepší je místo *Weight(m,n)* používat $..,m\$n,...$

Package describe - popisná statistika

$describe$ - popisná statistika

Všechny položky se volají pomocí

$with(stats)$; $describe[...](data)$; nebo zkráceně $stats[describe,....](data)$;

1. $count$

$describe[count](data)$;

uvede počet dat, položky $missing$ jsou ignorovány;

položky $Weight[..,n]$ nebo $...\$n$, se započítají jako n hodnot.

2. *range* rozpětí datového souboru;

$describe[range](data)$;

uvede maximální a minimální hodnoty ve tvaru $min..max$, položky $missing$ jsou ignorovány.

3. *countmissing*

$describe[countmissing](data);$

uvede počet chybějících položek v datech.

4. *sumdata* výběrový úhrn $\tilde{X}$; položky missing jsou ignorovány;

a) $describe[sumdata](data);$ uvede součet položek $\tilde{X} = \sum_{i=1}^{n} x_i;$

b) $describe[sumdata[k]](data);,\ k \in \boldsymbol{N}$

uvede součet $\sum_{i=1}^{n} x_i^k;$

c) $describe[sumdata[k, R]](data);\ \ \in \boldsymbol{N},\ R \in \boldsymbol{R}$

uvede součet $\sum_{i=1}^{n} (x_i - R)^k.$

5. *mode* modus

$describe[mode](data);$

uvede položku(y) s největší četností výskytu;

6. *median* medián

$describe[median](data);$

uvede medián $\tilde{x}$ souboru, kde pro lichý počet dat $\{x_1, x_2, \ldots, x_{2m+1}\}$ je $\tilde{x} = x_{m+1}$, pro sudý počet dat $\{x_1, x_2, \ldots, x_{2m}\}$ je $\tilde{x} = \frac{1}{2}(x_m + x_{m+1})$. Položky *missing* jsou ignorovány.

7. *mean* průměr (výběrový průměr);

$describe[mean](data);$

uvede průměr hodnot $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$, přičemž jsou položky *missing* ignorovány.

8. *meandeviation* průměrná odchylka;

$describe[meandeviation](data);$

uvede odchylku $d = \frac{1}{n} \sum_{i=1}^{n} |x_i - \overline{X}|.$

9. *quartile* kvartily

$describe[quartile[k]](data);\quad k \in \{1, 2, 3\}$, kde

$i = 1$ uvede dolní kvartil $x_{0,25}$;

$i = 2$ uvede medián $\tilde{x} = x_{0,5}$;

$i = 3$ uvede horní kvartil $x_{0,75}$.

Poznamenejme $IQR = x_{0,75} - x_{0,25}$ je *mezikvartilové rozpětí.*

10. *quantile* kvantily, kvantilová funkce;
$describe[quantile[\alpha]](data)$; $\alpha \in (0,1)$;
uvede $\alpha-$kvantil souboru;
$quantile[1/2]$ uvede medián a to nejbližší menší hodnotu ze souboru;
$quantile[1/2, 1/2]$ uvede průměr nejbližších hodnot k mediánu jako 6.

11. *decile* decily, $0, i-$kvantily, $i = 1, 2, \ldots, 9$;
$describe[decile[i]](data)$; $i = 1, 2, \ldots, 9$.
uvede decil $x_{0,i}$.

12. *percentile* percentily, $0, mn-$kvantily;
$describe[percentile[mn]](data)$; $m, n = 0, 1, 2, \ldots, 9$
uvede mn-percentil, tedy kvantil $x_{0,mn}$.

13. *coeficientofvariation* variační koeficient pro kladná data;
$describe[coeficientofvariation[i]](data)$; $i = 0, 1$;
$i = 0$ není třeba zadávat, je nutné pouze $i = 1$;
uvede $v = \frac{s}{\bar{x}}$ pro $i = 0$ a
$V = \frac{S}{\bar{x}}$ pro $i = 1$.

14. *standarddeviation* směrodatná odchylka
$describe[standarddeviation[i]](data)$; $i = 0, 1$;
hodnotu $i = 0$ není třeba zadávat, nutná je hodnota $i = 1$;
uvede s pro $i = 0$ a S pro $i = 1$.

15. *skewness* šikmost;
$describe[skewness](data)$;
uvede koeficient šikmosti A_3.

16. *kurtosis* špičatost;
$describe[kurtosis](data)$;
uvede koeficient špičatosti A_4, pro rozdělení $N(0,1)$ je $A_4 = 3$.

17. *moment* centrální a obecné výběrové momenty;
$describe[moment[k, R, j]](data)$;
$k \in \boldsymbol{N}$;
$R \in \boldsymbol{R}$, pokud se nezadá je $R = 0$;
$j = 0, 1$, pokud se nezadá je $j = 0$;
uvede moment $\frac{1}{n-j} \sum_{i=1}^{n} (x_i - R)^k$;
Jako parametr R je možné uvést *mean*.
$describe[moment[2, mean, 1]](data)$; uvede výběrový rozptyl S^2.

18. *variance* rozptyl, výběrový rozptyl
$describe[variance, i](data); \quad i \in \{0, 1\};$
pro $i = 0$ (nepovinné) uvede rozptyl s^2
pro $i = 1$ uvede výběrový rozptyl S^2, uvadí tedy $\frac{1}{n-i} \sum_{i=1}^{n} (x_i - \overline{X})^2.$

19. *covariance* kovariance $C(X, Y)$
$describe[covariance](data1, data2);$
uvede koeficient kovariance $\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}).$

20. *linearcorrelation* Pearsonův koeficient korelace
$describe[linearcorrelation](data1, data2)$
uvede $\frac{C(X,Y)}{S_X S_Y}.$

21. *quadraticmean* kvadratický průměr x_K
$describe[quadratimean](data);$
uvede $x_K, \ x_K^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2.$

22. *geometricmean* geometrický průměr x_G pro $x_i > 0$
$describe[geometricmean](data);$
uvede $x_G = \sqrt[n]{x_1.x_2 \ldots x_n}.$

23. *harmonicmean* harmonický průměr x_H pro $x_i > 0$
$describe[harmonicmean](data);$
uvede $x_H, \ x_H^{-1} = \frac{1}{n} \sum_{i=1}^{n} x_i^{-1}.$

Některé další příkazy

$map(f, data)$ nahradí položky $x[i]$ souboru data položkami $f(x[i])$, kde f musí být zadána jako funkce;

$sort(data)$ uspořádá číselná data od nejmenšího po největší do rostoucí posloupnosti, (pokud soubor neobsahuje data typu *missing* a *Weight*);

$convert(data, set)$ převede seznam na množinu (opakovaná data počítá jednou);

$convert(data, list)$ vytvoří z množiny opět seznam.

$unapply(v, x)$ vytvoří z výrazu v funkci $v(x)$