



StatSoft

# Rozdělení náhodné veličiny

V **předchozím dílu** našeho vzdělávacího seriálu jsme si ukázali nejzákladnější míry variability, které vhodně doplňují míry polohy v rámci popisné statistiky, ale také jsou základem pro celou řadu statistických testů. Nyní bychom se na rozdělení náhodných veličin podívali obecně.

Změříme-li nějakou veličinu, jde z hlediska teorie o náhodný pokus. Předpis, který přiřazuje každému výsledku našeho náhodného pokusu určité číslo, se nazývá náhodná veličina. Z matematického hlediska je tedy náhodná veličina  $X$  reálná funkce definovaná na množině všech elementárních jevů (jednotlivé možné výsledky pokusu), která každému jevu přiřadí reálné číslo. Pravděpodobnost, s kterou náhodná proměnná nabývá určité hodnoty nebo je obsažena v určitém intervalu hodnot se nazývá pravděpodobnostní rozdělení.

Pro vysvětlení principu začněme snad nejtypičtějším příkladem. Vpravo vidíte minci, pokud s touto mincí hodíte a sledujete, co padlo, provedete tím náhodný pokus.

Definičním oborem (možnými výsledky) tohoto pokusu jsou 2 možnosti:

- › první možnost = **padne líc** (panna)
- › druhá možnost = **padne rub** (orel)



Obor hodnot je množina  $\{0,1\}$ . Pro tuto situaci vlastně definiční obor a obor hodnot splývají, ale jak uvidíme v dalších příkladech, nemusí tomu tak být.

## Diskrétní náhodná veličina

Náhodné veličiny mohou být v zásadě dvojího typu – diskrétní a spojité. Již podle úvodního příkladku s mincí jistě vytušíte, že diskrétní náhodná veličina se vyznačuje tím, že může nabývat pouze spočetně izolovaných hodnot z nějaké množiny (vzpomeňte, o diskrétních veličinách jsme mluvili také v prvním díle našeho [vzdělávacího newsletteru](#)).

Každá z těchto hodnot  $\{x_1, x_2, \dots, x_n\}$  pak může být nabyta s určitou pravděpodobností  $\{p_1, p_2, \dots, p_n\}$ . Pravděpodobnostní rozdělení je tedy popsáno v tomto případě tabulkou vpravo. Každé hodnotě  $x_i$  je tedy přiřazena právě jedna hodnota  $p_i$ , přičemž platí, že součet  $p_i$  je roven jedné. To, že je součet právě jedna, nahlédneme jednoduše - pokud vypíšeme do tabulky všechny možné hodnoty, které může veličina nabývat, pak hodnota menší než jedna by znamenala, že může nastat i něco jiného než možnosti v tabulce, naopak hodnota větší než jedna by

|       |       |       |     |       |
|-------|-------|-------|-----|-------|
| $x_i$ | $x_1$ | $x_2$ | ... | $x_n$ |
| $p_i$ | $p_1$ | $p_2$ | ... | $p_n$ |

znamenal, že pravděpodobnost nabytí kterékoli z možností v tabulce je větší než jedna  $P(X \in \{x_1, x_2, \dots, x_n\}) > 1$ , což není možné.

Tabulka pro příklad s mincí by vypadala následovně, pokud tedy předpokládáme, že mince není „cinklá“.

|       |               |               |
|-------|---------------|---------------|
| $x_i$ | 0             | 1             |
| $p_i$ | $\frac{1}{2}$ | $\frac{1}{2}$ |

Rozdělení diskretních náhodných veličin lze popsat několika způsoby (všechny dávají stejnou plnou informaci o rozdělení):

1. Tabulkou pravděpodobností (řada rozdělení).

Takovouto tabulku jsme si již ukázali.

2. Pravděpodobnostní funkcí  $P(X = x_i) = p_i$ .

Pro všechny možnosti výsledku náhodné veličiny přiřazuje danou pravděpodobnost.

3. Distribuční funkcí  $F(x) = P(X \leq x)$ .

Ta tedy udává, jaká je pravděpodobnost, že výsledek náhodné veličiny bude menší nebo roven  $x$ .

## Spojité náhodná veličina

Spojité náhodné veličiny jsou takové veličiny, které mohou nabývat nekonečně hodnot. Příkladem takové veličiny může být například **doba čekání na nějakou událost** nebo **délka součástky**.

Rozdělení spojitě náhodné veličiny lze popsat následovně:

1. Hustotou  $f(x)$ .
2. Distribuční funkcí  $F(x) = P(X \leq x)$ .

Vztah mezi hustotou a distribuční funkcí je následující:  $F(x) = \int_{-\infty}^x f(t) dt$ .

## Příklady pravděpodobnostních rozdělení

Pomocí distribučních funkcí či hustot jsou v statistických publikacích popsány nejrůznější typy rozdělení náhodných veličin. Cílem tohoto článku není popisovat do detailu něco, co už bylo několikrát kvalitně popsáno, ukažme si proto spíše jen některé příklady pravděpodobnostních rozdělení a kde na ně můžeme v reálném životě narazit.

Pro základní přehled o rozděleních bychom například doporučili následující publikace: [Přehled statistických metod zpracování dat...](#) nebo [Základy matematické statistiky](#).

Pojďme se podívat na některé příklady použití rozdělení pro diskretní i spojitě veličiny, začneme diskretními veličinami.

# Diskrétní rozdělení náhodných veličin

## Alternativní rozdělení

Týká se náhodných veličin, které mohou mít pouze dva různé výsledky. Pravděpodobnost jednoho z nich je  $p$ , druhého  $1 - p$ . Značení tohoto rozdělení  $A(p)$ .

Příklad 1: Zmiňovaný **hod mincí** s  $p = 0,5$  a možnými výsledky  $\{0,1\}$ ,

Příklad 2: V pytlíku máme **tři kuličky černé** barvy a **jednu bílé barvy**, výsledkem pokusu bude vytažení černé či bílé kuličky, bílé s pravděpodobností  $p = 0,25$  a černé s pravděpodobností  $1 - p = 0,75$ .

## Rovnoměrné rozdělení

Je  $n$  možných výsledků a všechny jsou stejně pravděpodobné, tedy  $p_i = \frac{1}{n}$ . Značení  $R(n)$ .

Příklad 1: **Hod kostkou**. Pravděpodobnost, že padne 1 je stejná jako pravděpodobnost, že padne 2, atd.

|       |               |               |               |               |               |               |
|-------|---------------|---------------|---------------|---------------|---------------|---------------|
| $x_i$ | 1             | 2             | 3             | 4             | 5             | 6             |
| $p_i$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ |



Příklad 2a: **Ruleta (evropská)**. Pravděpodobnost, že se kulička zastaví na daném čísle z rozmezí od 0 do 36 je  $p_i = \frac{1}{37}$ .

Příklad 2b: **Ruleta (americká)**. Zde je pravděpodobnost  $p_i = \frac{1}{38}$ . Divíte se, proč je to jinak než v předchozím příkladu? Není se čemu divit, na americké ruletě je totiž navíc kromě čísel od 0 do 36 ještě číslo 00.

## Binomické rozdělení

Veličina s binomickým rozdělením je součtem nezávislých veličin s rozdělením alternativním, parametry tohoto rozdělení jsou tedy jednak  $p$  a poté ještě počet pokusů  $n$ . Toto rozdělení se typicky značí jako  $Bi(n, p)$ . Pravděpodobnostní funkce vypadá takto

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x},$$

kde  $x$  může nabývat hodnot 0, 1, 2 až  $n$ .

Příklad 1: **Počet šestek při hodu 3 kostkami**. Například pravděpodobnost, že padne jedna šestka je  $P(X = 1) = \binom{3}{1} \frac{1}{6} \left(\frac{5}{6}\right)^2 = \frac{25}{72} = 0,347$ .



Otázka: K tomuto příkladu by nás mohla napadnout například otázka: Co má větší pravděpodobnost? Že hodíme třemi kostkami jednu šestku nebo šesti kostkami dvě šestky? Co myslíte?

Odpověď: Pro tři kostky a jednu šestku již výpočet máme, stačí tedy provést ještě výpočet binomického rozdělení pro  $n = 6$  a  $p = \frac{1}{6}$ . Tedy  $P(Y = 1) = \binom{6}{2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^4 = 0,2$ . Pravděpodobnost hození jedné šestky třemi kostkami je tedy větší.

Příklad 2: **Hod jednou kostkou.** Sledujeme, zda padne šestka. Toto jsme zde již měli, toto je pokus s rozdělením  $A\left(\frac{1}{6}\right)$ , což je vlastně  $Bi\left(1, \frac{1}{6}\right)$ .

Poznámka: Binomické rozdělení je například ten případ, kdy nemáme stejný definiční obor a obor hodnot u náhodné veličiny. Vezměme příklad šesti kostek: definiční obor jsou všechny výsledky hodu (například padlo: 1, 2, 2, 4, 6, 6). Nicméně oborem hodnot je pouze počet šestek.

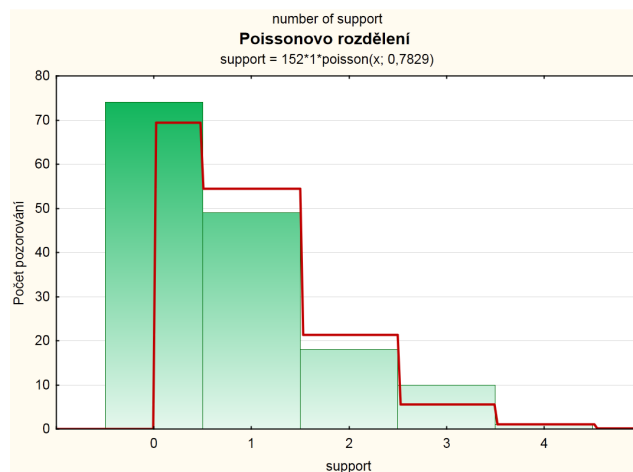
## Poissonovo rozdělení

Nabývá kladných celých čísel včetně 0 a většinou vyjadřuje počet výskytů málo pravděpodobných jevů v nějakém časovém či objemovém intervalu. Toto diskrétní rozdělení také vznikne jako limitní případ binomického rozdělení pro velká  $n$ . Poissonovo rozdělení má jeden parametr, většinou značený jako  $\lambda$ , který je zároveň střední hodnotou i rozptylem. Rozdělení značíme jako  $Po(\lambda)$ . Pravděpodobnostní funkce pak vypadá takto

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}.$$

Příklady: **Počet úrazů za určitý interval, počet tiskových chyb na jedné stránce, počet branek za fotbalový zápas, počet vadných výrobků za časový interval,** atd.

Příklad z naší kuchyně: **Počet žádostí o technickou podporu za den.** Vyzkoušejme poissonovo rozdělení na reálná data z naší firmy, vytáhli jsme si z databáze počty žádostí o technickou podporu za poslední přibližně půlrok, histogram vidíte na obrázku. Spolu s histogramem je vidět i proložení odhadnutým poissonovým rozdělením. Sami můžete posoudit, jak se k sobě skutečná data a teoretické rozdělení blíží.



## Geometrické rozdělení

Geometrické rozdělení je definováno jako počet úspěchů před prvním neúspěchem. Všechny pokusy musí být nezávislé a mít konstantní pravděpodobnost úspěchu  $p$ . Rozdělení  $Ge(p)$  je definováno jako

$$P(X = x) = p^k(1 - p).$$

Příklad 1: **Hod mincí.** Jaká je pravděpodobnost, že hodíme pannu až v pátém hodu?

$$P(X = 4) = \left(\frac{1}{2}\right)^4 \left(1 - \frac{1}{2}\right) = 0,03125.$$

## Negativně binomické rozdělení

Toto rozdělení je zobecněním geometrického rozdělení.  $NBi(r, p)$  modeluje počet úspěchů před  $r$ -tým neúspěchem.

$$P(X = x) = \binom{r+k-1}{k} p^r(1-p)^k.$$

## Spojitá rozdělení náhodných veličin

### Rovnoměrné rozdělení

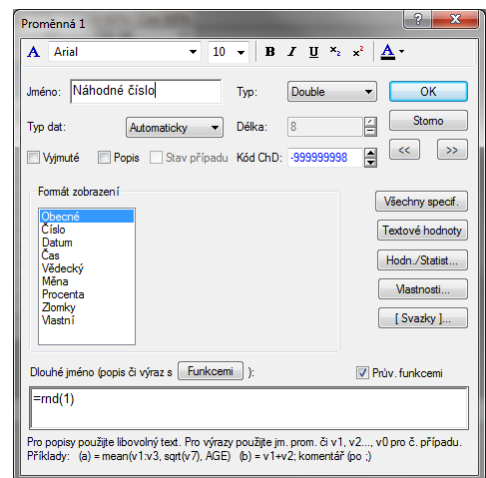
Stejně jako pro diskrétní rovnoměrné rozdělení, i zde platí nepreferovanost některých hodnot nad jinými. Hustota rovnoměrného rozdělení  $R(a, b)$  na intervalu  $(a, b)$  je konstantní

$$f(x) = \frac{1}{b-a} \text{ na intervalu } x \in (a, b), \text{ jinak } f(x) = 0.$$

Příklad 1: Každý správný **generátor náhodných čísel** by měl být schopen generovat z rovnoměrného rozdělení na intervalu  $(0,1)$ . Ve *STATISTICE* nagenerujeme náhodná čísla například pomocí funkce *Rnd* (viz obrázek vpravo).

Příklad 2: **Zaokrouhlovací chyba.** Pokud zaokrouhlíme reálné číslo na celé číslo, pak zaokrouhlovací chyba bude mít rozdělení  $R\left(-\frac{1}{2}, \frac{1}{2}\right)$ .

Příklad 3: **Doba čekání na metro.** V náhodný okamžik vejde do stanice metra, víme, že interval mezi příjezdy jednotlivých vlaků je 8 minut. Doba čekání na metro bude mít tedy rovnoměrné spojitě rozdělení  $R(0,8)$ .



### Exponenciální rozdělení

Exponenciální rozdělení  $Exp(\lambda)$  má hustotu

$$f(x) = \lambda e^{-\lambda x} \text{ pro } x \geq 0, \text{ jinak } f(x) = 0.$$

Typicky vyjadřuje čas mezi náhodně se vyskytujícími se událostmi.

Příklady: [Doba čekání mezi poruchami přístroje](#), [doba životnosti žárovky](#), [aplikace v teorii hromadné obsluhy – doba čekání ve frontě](#), atd.

Poznámka: Exponenciální rozdělení má speciální vlastnost, je to takzvané rozdělení „bez paměti“. Demonstrujme si to na příkladu žárovky, pokud se doba životnosti žárovky chová podle exponenciálního rozdělení, pak to, že víme, že žárovka již nějakou dobu svítí, nám nepřinese žádnou informaci o tom, jak dlouho ještě bude svítit – pravděpodobnost je stejná jako bychom žárovku rozsvítili poprvé.

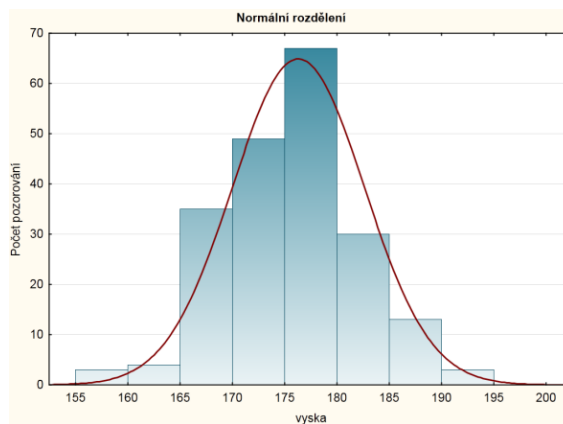
## Normální rozdělení

A konečně se dostáváme k neznámějšímu a snad nejdůležitějšímu rozdělení. Označováno také jako **Gaussovo rozdělení**. Důležité je především z důvodu platnosti centrální limitní věty, která říká, že součet nezávislých náhodných veličin má přibližně normální rozdělení (čím více veličin sečteme, tím lépe), což znamená, že lze aproximativně použít procedury, které byly na základě normálního rozdělení navrženy, pro proměnné, které se tímto rozdělením neřídí. Toto rozdělení se objevuje u jevů, jejichž proměnlivost je způsobena velkým množstvím nezávislých veličin. Normální rozdělení  $N(\mu, \sigma^2)$  má hustotu

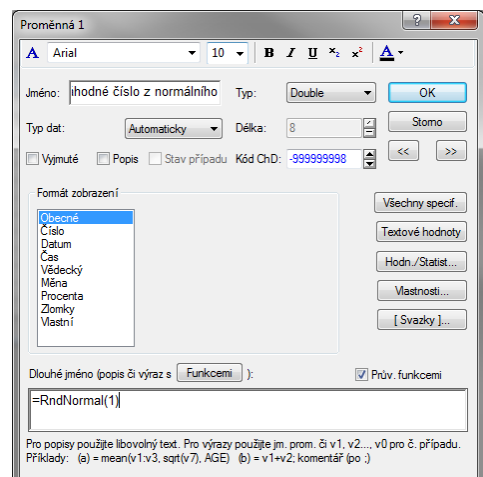
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}.$$

Parametry  $\mu$  a  $\sigma$  jsou střední hodnota a směrodatná odchylka rozdělení, těmito dvěma parametry je již normální rozdělení určeno jednoznačně.

Příklady: Typicky se normálním rozdělením popisuje například [výška osob v populaci](#), [tělesná teplota](#), [výnosy těžé plodiny na různých polích](#), [chyby měření](#), [rychlost molekul v plynu](#), atd.



Normální rozdělení  $N(0,1)$  nazýváme normované normální rozdělení. Náhodné hodnoty z tohoto rozdělení je možné v programu *STATISTICA* mimo jiné nagenarovat přes funkci *RndNormal*.



## Rozdělení vycházející z normálního rozdělení

Na závěr bychom si představili rozdělení, která z normálního rozdělení vycházejí a se kterými se setkáváme hlavně v teorii testování hypotéz. Těmito rozděleními jsou:

### Chí-kvadrát rozdělení

$\chi^2(n)$  je rozdělení veličiny, která vznikne jako  $\sum_{i=1}^n U_i^2$ , kde  $U_i$  jsou nezávislé náhodné veličiny s rozdělením  $N(0,1)$ .

Poznámka: Chí-kvadrát rozdělení známe například díky aplikaci chí-kvadrát testu. Nicméně skrze Studentovo a F rozdělení, které jsou na chí-kvadrát rozdělení založeny, se promítá i do většiny dalších parametrických testů.

### Studentovo t rozdělení

Studentovo rozdělení o  $n$  stupních volnosti  $t(n)$  je rozdělení náhodné veličiny  $\frac{U}{\sqrt{V/n}}$ , kde  $U$  má  $N(0,1)$  rozdělení a  $V$  má  $\chi^2(n)$  rozdělení,  $U$  a  $V$  jsou nezávislé.

Poznámka: Studentovo rozdělení je využíváno například v testech na shodnosti středních hodnot nebo v regresi při testech nulovosti regresních koeficientů.

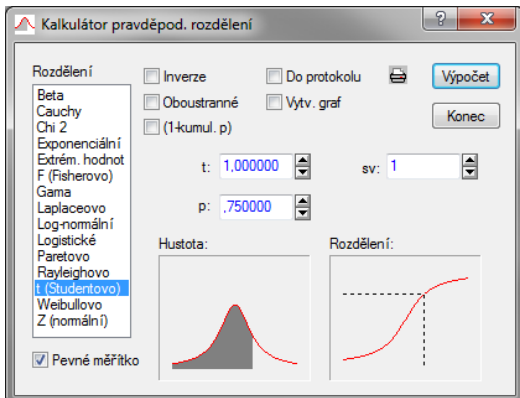
### Fisherovo-Snedecorovo rozdělení (nebo také F rozdělení)

F rozdělení vzniká jako rozdělení podílu dvou nezávislých náhodných veličin s chí kvadrát rozdělením, tedy veličina s rozdělením  $F(m, n)$  má stejné rozdělení jako veličina  $\frac{U/m}{V/n}$ , kde  $U$  a  $V$  jsou nezávislé,  $U$  má  $\chi^2(m)$  a  $V$  má  $\chi^2(n)$  rozdělení.

Poznámka: F rozdělení je například využíváno v analýze rozptylu.

Pokud se Vám zdá, že jsme již článek příliš natáhli, nenechte se mýlit, téma pravděpodobnostních rozdělení je mnohem širší než by se mohlo zdát. Jako důkaz přijměte ochutnávku knih profesora [Normana Lloyda Johnsona: Univariate Discrete Distributions, Continuous Univariate Distributions 1](#) a [Continuous Univariate Distributions 2](#). Každá z těchto knih má okolo 700 stran a to se jedná pouze o jednorozměrná rozdělení.

# Pravděpodobnostní kalkulátor

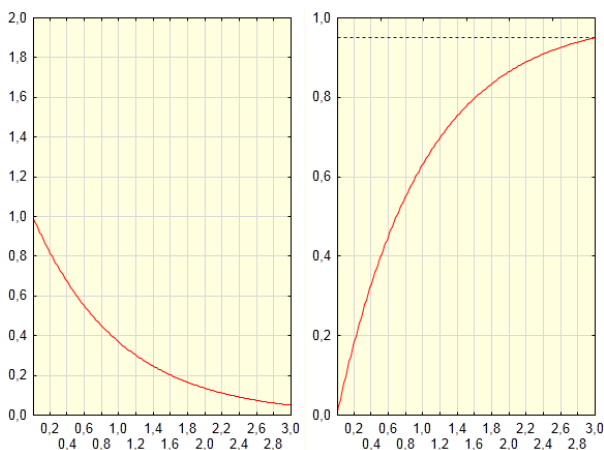


Pro výpočet hodnot distribučních funkcí, hustot či kvantilových funkcí spojitých rozdělení můžete využít **pravděpodobnostní kalkulátor** v softwaru *STATISTICA*. V softwaru jej najdete v menu *Statistiky* -> *Pravděpodobnostní kalkulátor* -> *Rozdělení...*

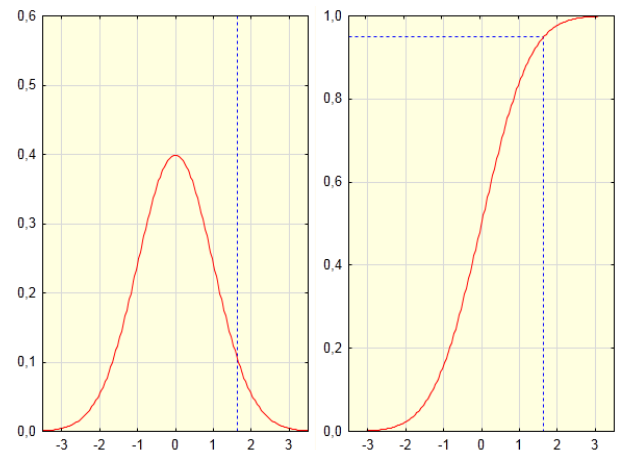
Zaškrtnutím volby *Výtv. Graf* a potvrzením tlačítkem *Výpočet* můžete vytvořit graf s křivkami v okénkách dole.

Pro srovnání jednotlivých rozdělení jsme vykreslili některá rozdělení v tomto nástroji – viz obrázky níže, jedná se o hustoty a distribuční funkce příslušných rozdělení. Čárkované je zaznačen 95 procentní kvantil.

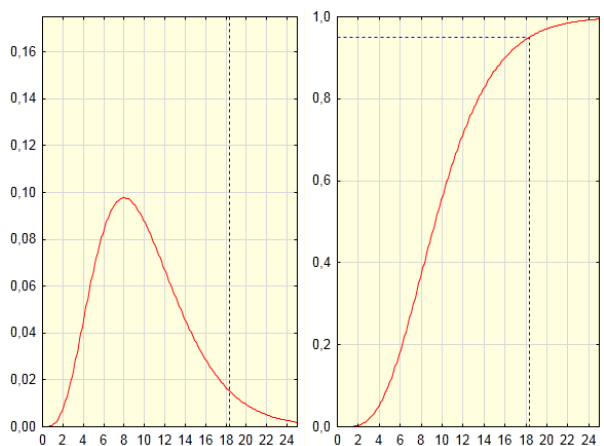
$Exp(1)$



$N(0,1)$



$\chi^2(10)$



$t(1)$

