## Náhoda při hře


Martingale:

- Vsadíš řekněme 1 dolar na barvu, kterou si vybereš (červená či černá) a budeš stále sázet jen na ni.
- Roztočíš ruletu a čekáš …
- Pokud prohraješ, zdvojnásobíš sázku, takže vsadíš příště 2 dolary. Pak řekněme, že znova prohraješ, tak příště vsadíš 4 dolary. A když zas prohraješ, tak příště vsadíš 8 dolarů. Řekněme, že už konečně padla barva, kterou jsi vsadil. Casino ti dá ke tvým vsazeným 8 dolarům svých 8, takže máš teď 16 dolarů. Spočítej si, kolik jsi vsadil: 1 + 2 + 4 + 8 = 15 dolarů, takže - vydělal jsi 1 dolar!
- Pak začneš sázet znova 1 dolar…

## Náhoda při hře


Pravděpodobnost výhry při sázce na barvu:

$p = 18/37 = 0{,}486$

Průměrný “zisk” při *n* sázkách částky č:

$- n.č + 2.č.n.p = n.č.(-0{,}027)$

Dlouhodobý zisk kasina je 2,7% ze vsazených částek.

## Co to je pravděpodobnost?

- je matematická disciplína, popisující zákonitosti týkající se jevů, které (přinejmenším z hlediska pozorovatele) mohou a nemusí nastat, resp. jejichž výsledná hodnota není předem jistá.
- Příkladem může být výsledek hodu kostkou ještě předtím, než hodíme, anebo venkovní teplota zítra v poledne.

- **Laplaceova def. pravděpodobnosti**
  Může-li náhodný pokus vykázat konečný počet *n* různých, vzájemně se vylučujících, stejně možných výsledků a jestli *m* z těchto výsledků má za následek událost *A*, pak pravděpodobnost události A položíme rovnou číslu *m/n*

$$P(A) = \frac{m}{n}$$

## Model pro hod kostkou

Úkol: Odhadněte pst, že padne šestka

100 x opakujeme pokus

`hody=ceil(6*rand(100,1))` transformace U(0,1)

nebo

`hody2=randi(6,100,1)` diskrétní U(1, 2, …6)

Bodovým odhadem pravděpodobnosti je relativní četnost.

`pst6=sum(hody==6)/1000`

# Základní pojmy

**Náhodný pokus** -výsledky závisí nejen na předepsaných podmínkách, ale také na **náhodě**.

**Náhodný jev** - každý výsledek (elementární jev) nebo skupina výsledků náhodného pokusu.
Jakékoliv tvrzení o výsledku náhodného pokusu, o kterém lze (po provedení pokusu) rozhodnout, zda je pravdivé.

**Prostor elementárních jevů *S*** - množina všech možných výsledků pokusů

– jsme schopni je všechny předem vyjmenovat *(ve sš matematice konečná),*

– jsou vzájemně neslučitelné *(žádné dva nemohou nastat současně),*

– jeden z nic nastane vždy *(nemůže nastat žádný jiný než jeden z jmenovaných),*

# Chevalier de Mere

- **Hod jednou kostkou**
  Chevalier de Mere příjimal sázky na to, že hodí minimálně jednu šestku ve čtyřech po sobě následujících hodech.
  *Pravděpodobnost padnutí šestky je v každém hodu 1/6. Domníval se, že jeho šance na padnutí šestky ve čtyřech hodech je tedy (1/6) × 4 = 2/3.*
- **Správně:** Vypočítáme pravděpodobnost doplňkového jevu k počtu nepříznivých možností: $P(A) = 1 - (5/6)^4 = 0{,}517746914$

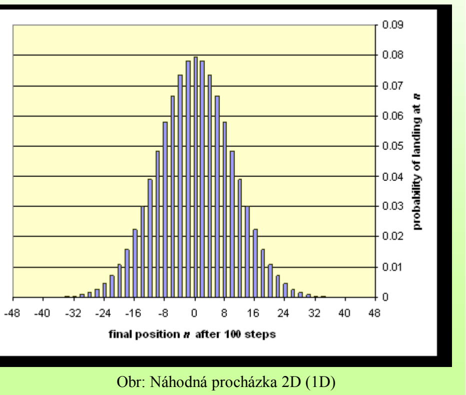
- **Hod dvěma kostkami.**
  Vyhrajeme, pokud se nám alespoň jedenkrát podaří hodit 2 šestky ve 24 hodech.
  *Pravděpodobnost vrhnutí dvou šestek v jednom hodu je 1/36.*
  *De Mere předpokládal, že jeho šance je tedy (1/36) × 24 a tedy opět 2/3 .*
- **Správně**: Pravděpodobnost výhry bude:
  $P(A) = 1 - (35/36)^{24} = 0{,}491403876$.

# Metoda Monte Carlo

- Metoda, která používá stochastických metod k řešení deterministických problémů.

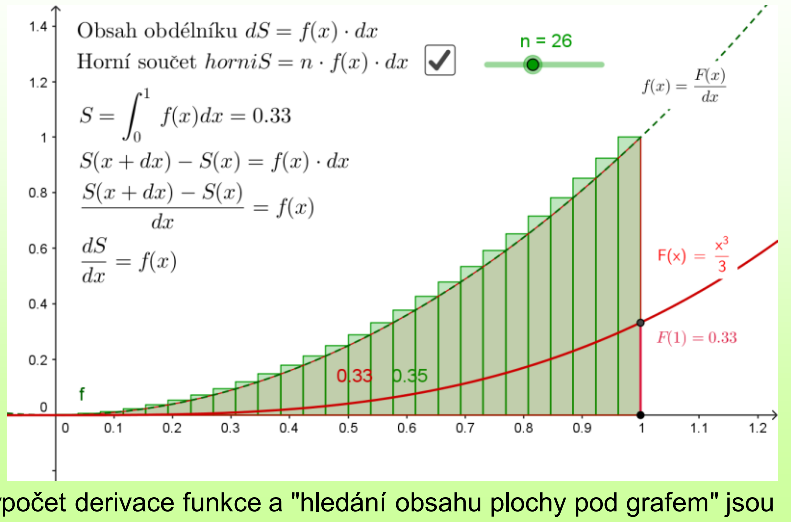
Postup:

1. Systém nahradíme simulačním modelem se stejnými pravděpodobnostními charakteristikami a chování mnohonásobně simulujeme na modelu.
2. Jednotlivé charakteristiky výstupu nahradíme bodovým odhadem
   - střední hodnotu průměrem
   - pst stavu relativní četností



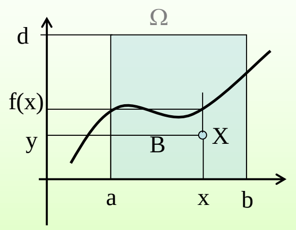

Obr: Náhodná procházka 2D (1D)

# Výpočet integrálu metodou Monte Carlo




Výpočet derivace funkce a "hledání obsahu plochy pod grafem" jsou "obrácené" operace.

# Výpočet integrálu metodou Monte Carlo




$$I = \int_a^b f(x)$$

$$\Omega = \langle a,b \rangle \times \langle 0,d \rangle$$

$$I = S_B$$

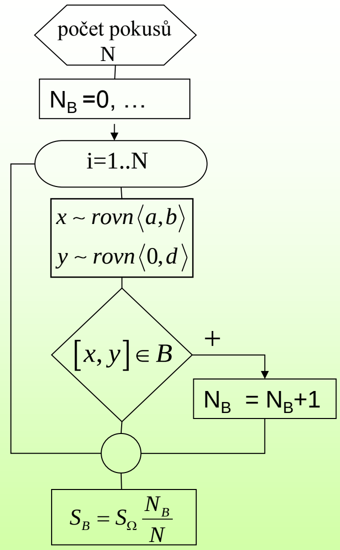
X … náhodně vybraný bod z Ω

$$X \in B \Leftrightarrow f(x) > y$$

$$P(X \in B) = \frac{S_B}{S_\Omega}$$

Pravděpodobnost odhadneme relativní četností

$$P(X \in B) = \frac{N_B}{N_\Omega}$$




# Simulace Monte Carlo

```
%Určete obsah plochy y=x^2 pod
grafem na intervalu (0,1)
clc;clear;
datax=rand(100,1);
datay=rand(100,1);
NoS=sum(datay < datax.^2)
area=NoS/100
```

```
n=100;pokusy=500
area = zeros(1,n);
for i=1:n
datax=rand(pokusy,1);
datay=rand(pokusy,1);
NoS=sum(datay < datax.^2);
area(i)=NoS/pokusy;
end
mean(area)
```

https://scratch.mit.edu/projects/148016090/

## Metoda Monte Carlo

- Při každém běhu dostaneme jiný výsledek
  - Odpovídá reálné situaci
- Každý výsledek jednoho běhu stochastického simulačního modelu musí být považován za jedno pozorování statistického experimentu !
- Kolikrát musíme opakovat simulaci abychom mohli „důvěřovat“ výsledku?

## Intervalový odhad zkoumaných veličin

- Provádíme *n* nezávislých pokusů – získáme realizace *Xi* zkoumané náhodné veličiny . Interval spolehlivosti pro odhad střední hodnoty náhodné veličiny

$$\left( \bar{X} - t_{\frac{(1-\alpha)}{2},(n-1)} \frac{\sigma}{\sqrt{n}} ; \bar{X} + t_{\frac{(1-\alpha)}{2},(n-1)} \frac{\sigma}{\sqrt{n}} \right)$$

-  σ - standardní odchylka získaná ze vzorku
- α – hladina významnosti

$t_{\frac{(1-\alpha)}{2},(n-1)}$ kvantil Studentova rozdělení s (n-1) stupni volnosti

# Statistika

aneb známe tři druhy lži:
- úmyslná
- neúmyslná
- statistika

Statistika je metoda, jak vyjádřit nejistá data s přesností na setinu procenta.

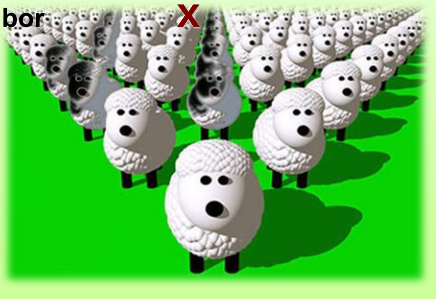
# Statistika

- Shromažďování a třídění dat
- Analýza za účelem formulování obecných závěrů a rozhodování

Popisná statistika - Metody zjišťování a sumarizace informací
Inferenční statistika - Měření spolehlivosti závěrů o populaci založených na informacích získaných z výběru

náhodný výběr

**výběrový soubor** X **základní soubor**



| den | teplota |
|---|---|
| 1.4.2008 | 11 |
| 2.4.2008 | 10 |
| 3.4.2008 | 10 |
| 4.4.2008 | 9 |
| 5.4.2008 | 8 |
| 6.4.2008 | 7 |
| 7.4.2008 | 8 |
| 8.4.2008 | 9 |
| 9.4.2008 | 4 |
| 10.4.2008 | 9 |
| 11.4.2008 | 8 |
| 12.4.2008 | 7 |
| 13.4.2008 | 8 |
| 14.4.2008 | 9 |
| 15.4.2008 | 12 |
| 16.4.2008 | 13 |
| 17.4.2008 | 15 |
| 18.4.2008 | 11 |
| 19.4.2008 | 12 |
| 20.4.2008 | 10 |
| 21.4.2008 | 9 |
| 22.4.2008 | 8 |
| 23.4.2008 | 9 |
| 24.4.2008 | 11 |
| 25.4.2008 | 10 |
| 26.4.2008 | 9 |
| 27.4.2008 | 6 |

# Elementární zpracování dat

- **cílem je zjednodušit nějaká data tak, abychom se v nich lépe vyznali**
- **důsledkem je ztráta informací!**

*průměrná teplota: 9.2°C*
*minimum: 4°C*
*maximum: 15°C*
*rozsah: 11°C*
*modus: 9°C*
*medián: 9°C*
*rozptyl: 5.1°C*
*směrodatná odchylka: 2.3°C*

# Hodnota statistického znaku

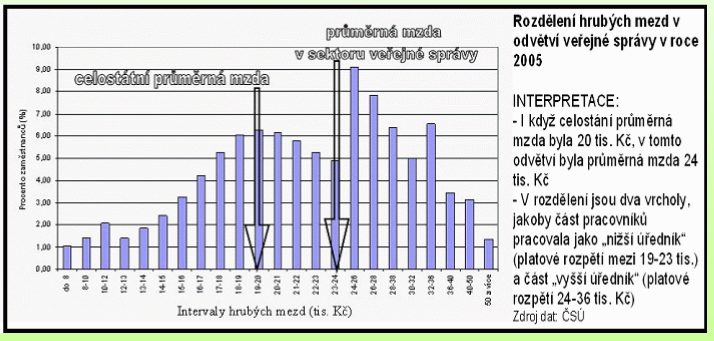
1. **Kvalitativní**
   - Nominální – kvalitativní znaky
     Jednotlivé hodnoty jsou neporovnatelné, data nelze seřadit, neexistuje nic jako „velikost“
   - Ordinární – lze je seřadit a přiřadit jim číslo, tj. určit pořadí
     Neumožňují posoudit vzdálenost
2. **Kvantitativní**
   - Diskrétní (je možné je spočítat)
   - Spojité (údaje, které lze měřit -přečíst na stupnici)

Určete typ dat:

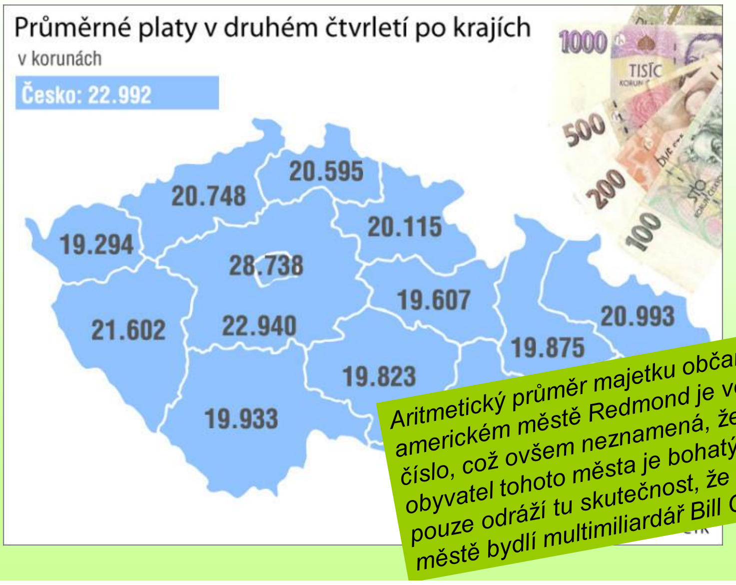
věk člověka, výška, počet podlaží, teplota, pohlaví, zaměstnání, účel budovy, barva očí, dosažené vzdělání, počet studentů, známka ze zkoušky,

# **Třídění** – rozdělení dat do tříd

1. Tříd musí být „tak akorát“
   *cílem je potlačit kolísání četností, ale nesmíme setřít charakteristické rysy*
2. Třídy musí být disjunktní
3. Stejná šířka intervalu




# Základní popisné statistiky




*Aritmetický průměr majetku občanů v americkém městě Redmond je velice vysoké číslo, což ovšem neznamená, že typický obyvatel tohoto města je bohatý. Tento fakt pouze odráží tu skutečnost, že v daném městě bydlí multimiliardář Bill Gates."*

## Charakteristiky polohy

- **aritmetický průměr** `mean(data)`
  výběrový průměr je nejlepší nestranný konzistentní odhad střední hodnoty základního souboru.
  Citlivý na outliery - vychýlené hodnoty můžeme odfiltrovat pomocí percentilu apod. (86% percentil říká, že 86% prvků leží pod touto hodnotou a 14% nad ní)

- **modus**
  - Nejčastěji se vyskytující hodnota `mode(data)`
  - min. modus = 1, max. modus = N
  - může jich být víc
  - odpovídá vrcholu histogramu četností
- **medián (50** % kvantil**)** `median(data)`
  - polovina pozorování menší než medián, polovina větší
  - střed uspořádaného základního souboru
  - i pro pouze „seřazená" data (na ordinální stupnici) – např. jídlo je vynikající (1), dobré (2), ucházející (3), bez chuti (4), nic moc (5), hnusné (6), vyvolávající zvracení (7)
  - v případ „ulítlé" hodnoty lepší vypovídající hodnota než průměr – robustní charakteristika

## Míry rozptýlení

$$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$$

- **rozptyl** (variance) `var(data)`
  - průměrná hodnota druhé mocniny odchylky od průměru

- **směrodatná odchylka** `std(data)`
  - odmocnina z rozptylu
  - čím menší, tím nižší variabilita dat

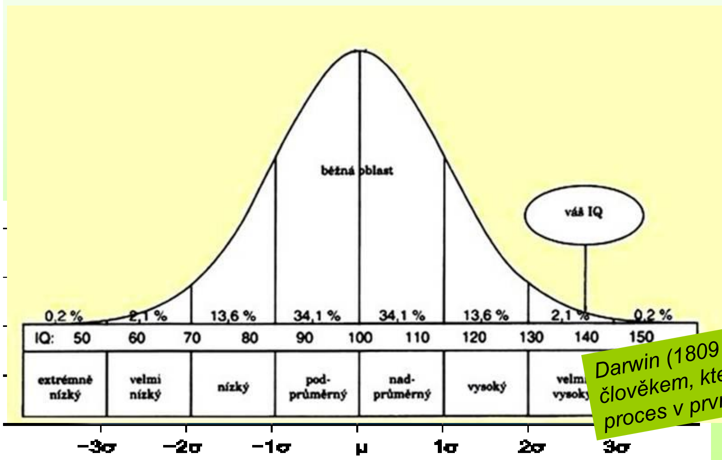
- **variační koeficient** – porovnává variabilitu nestejně velkých objektů (myš a slon) – bezrozměrné číslo

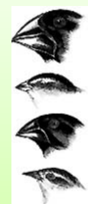
$$CV = \frac{\sigma}{\overline{X}}$$

- Qvartilové rozpětí
  | kvartil 75 % - kvartil 25 %|

- směrodatná odchylka
  - empirické pravidlo: většina hodnot se neodlišuje od průměru o více než jednu směrodatnou odchylku a skoro všechny hodnoty jsou v pásmu do dvou směrodatných odchylek od průměru.

normální rozdělení:



Darwin (1809 - 1882) byl prvním člověkem, který evoluci vnímal jako proces v první řadě statistický.

# Výpočetní složitost algoritmů

Složitost algoritmu udává, jak je daný algoritmus rychlý (kolik provede elementárních operací) vzhledem k množině vstupních dat.

**Asymtotická složitost** - rozdělení algoritmů do tříd složitostí, u kterých platí, že od určité velikosti dat, je algoritmus dané třídy vždy pomalejší než algoritmus třídy předchozí, bez ohledu na to, jestli je některý z počítačů výkonnější.

$$1 \ll \log(n) \ll n \ll n \cdot \log(n) \ll n^k \ll k^n \ll n! \ll n^n$$

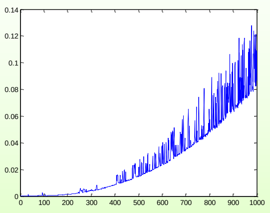
$O(f(n))$ znamená, že
$algoritmus\ probíhá\ asyptoticky\ stejně\ rychle\ \ nebo\ rychleji\ než\ f(n)$

Pro počet operací $g(n)$ platí
$\exists\, n_0,\ c;\ \forall n > n_0$ je $c \cdot f(n) < g(n)$

```
n=1:1:10;
plot(n,n,'r',n,n.*log(n),'g', 'LineWidth',2)
figure
plot(n, n.^2,'m',n, 2.^n,'c')
```

# Násobení matic

```
for i=10:1000;
tic;
soucin(i);
time(i) = toc;
 end
plot(time);
```

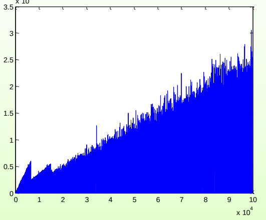


- `function[out]=soucin(n)`
- `%soucin dvou nahodne vygenerovanych matic nxn`
- `out = rand(n,n)*rand(n,n);`

soucintictoc.m

# Merge sort ?

```
time=zeros(1,100)
for n=1:100:50000
    L=randi(50,n,1);
    tic
    sort(L);
    time(n)=toc;
end
plot(time)
```

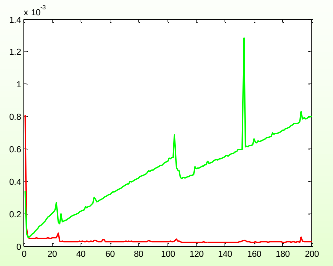


- A + B* n*log(n)

vyp_slozitost.m

# Faktoriál

https://cs.wikipedia.org/wiki/Rekurze_%28programov%C3%A1n%C3%AD%29




```
clear;
for n=1:200
    tic
    factorial(n);
    time1(n)=toc;
    tic
    factorial2(n);
    time2(n)=toc;
end
plot(1:200,time1,'r',1:200,time2,'g','LineWidth',2)
```

```
function[out]=factorial2(n)
if n==0
    out=1;
else out=factorial2(n-1)*n;
end
```

vyp_slozitost2.m