

Pravděpodobnostní metody a matematická statistika

Prof. RNDr. Gejza Dohnal, CSc.



Pravděpodobnostní metody a matematická statistika

- I. Úvod a motivace
- II. Pravděpodobnost (měření naděje)
- III. Náhodná veličina
- IV. Pravděpodobnostní modely
- V. Náhodný vektor
- VI. Náhodný proces
- VII. Časová řada
- VIII. Statistická inference, analýza dat
- IX. Odhadování parametrů
- X. Testování hypotéz
- XI. Analýza rozptylu
- XII. Regresní analýza
- XIII. Snižování dimenze



Pravděpodobnostní metody a matematická statistika

- (1) Jiří Likeš, Josef Machek: Počet pravděpodobnosti. Sešit X edice MVŠT, SNTL Praha 1981, 2. vydání 2019
- (2) Jiří Likeš, Josef Machek: Matematická statistika. Sešit XI edice MVŠT, SNTL Praha 1988, 2. vydání 2019
- (3) další učebnice naleznete na adrese: <https://sms.nipax.cz/pmvs>



I. Úvod a motivace

- **Variabilita**

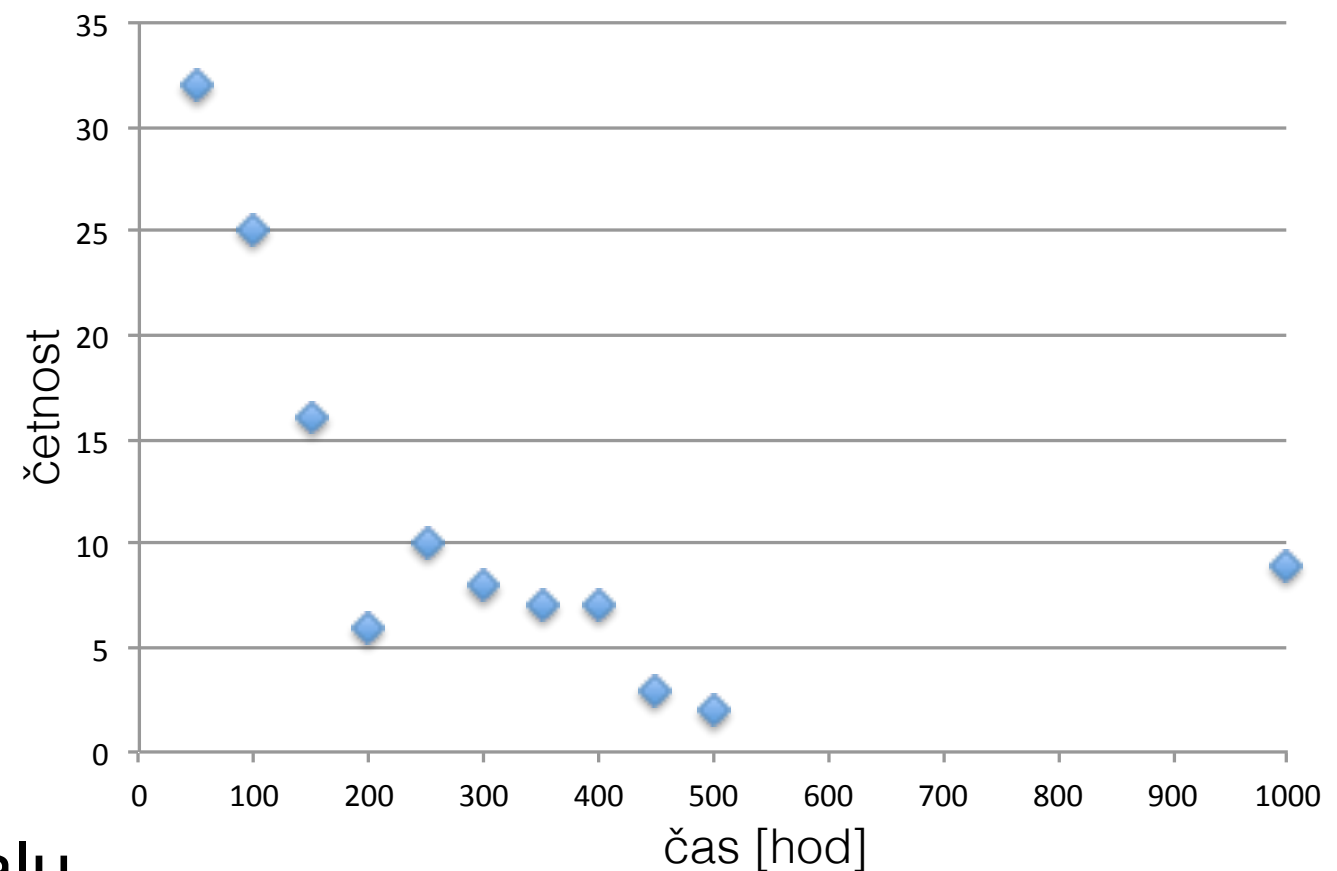


I. Variabilita

i	t_i	n_i
1	50	32
2	100	25
3	150	16
4	200	6
5	250	10
6	300	8
7	350	7
8	400	7
9	450	3
10	500	2
11	1000	9

i ... pořadové číslo intervalu,
 t_i ... horní mez intervalu i [hod],
 n_i ... četnost poruch v intervalu i ,

Doba do poruchy navigačních přístrojů



I. Variabilita

Doby mezi průjezdy automobilů mýtnou branou

0.26	5.00	8.65	7.20	1.62	2.91	0.17	6.05	2.70	3.95
1.12	3.18	3.27	7.45	3.08	0.81	3.65	0.40	0.32	0.23
0.57	6.24	4.65	6.49	2.80	5.62	11.78	4.17	0.48	1.15
0.93	1.32	3.43	0.28	5.95	0.63	1.91	3.94	2.57	4.81
1.81	5.04	2.05	12.64	1.73	10.54	5.75	5.41	5.21	2.39



$$Y_{\min} = 0.17$$

$$Y_{\max} = 12.64$$

$$Y_{\text{med}} = 3.13$$

$$Y_{\text{LQ}} = 1.15$$

$$Y_{\text{UQ}} = 5.41$$

0	2	2	3	3	3	4	5	6	6	8	9	1	1	3	6	7	8	9
2	1	4	6	7	8	9	1	2	3	4	7	9						
4	0	2	7	8	0	0	2	4	6	8								
6	0	1	2	5	2	5												
8	7																	
10	5	8																
12	6																	

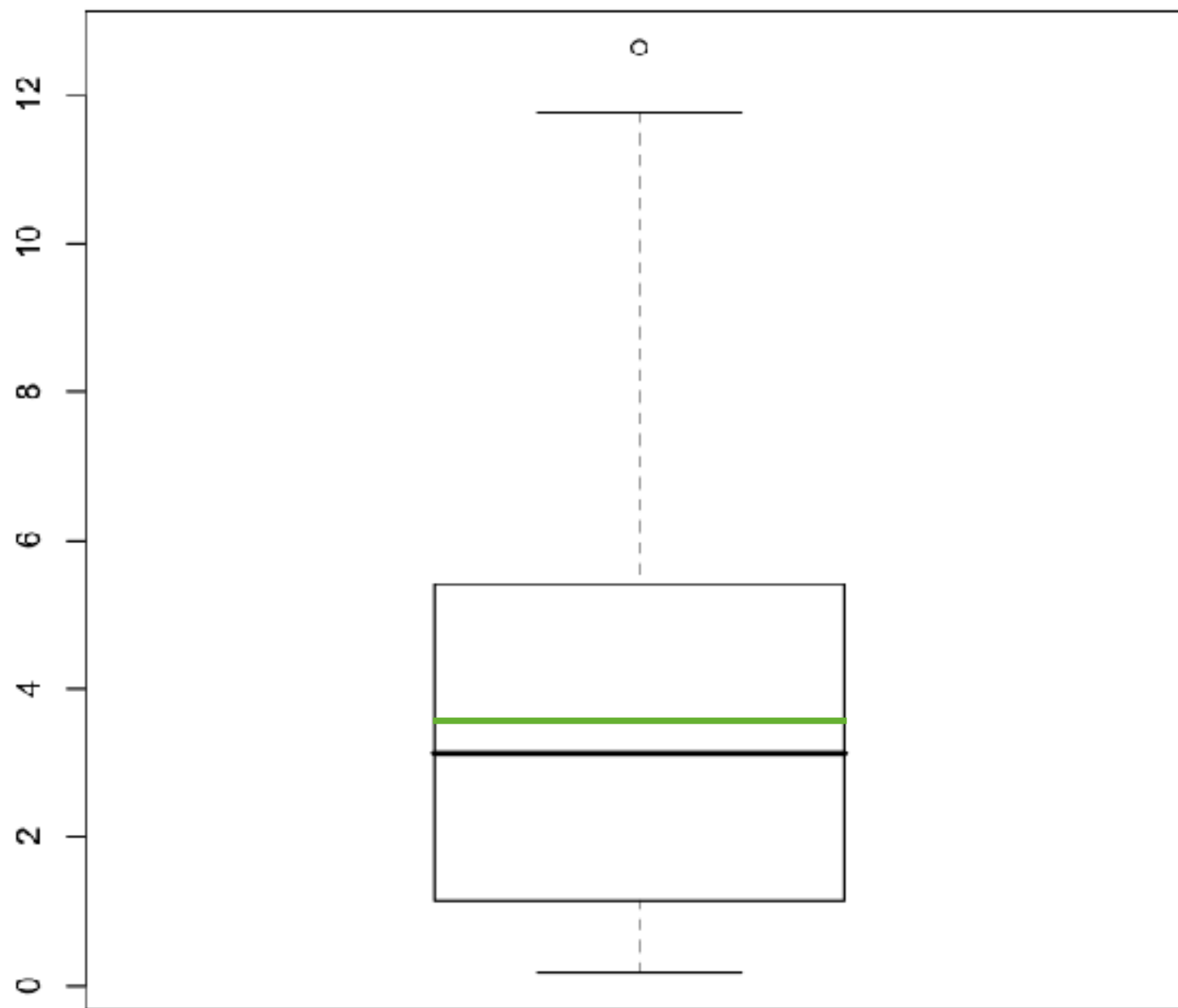
$$\bar{Y} = 3.69$$

$$s^2_{n-1}(Y) = 9.0957$$



I. Variabilita

Doby mezi průjezdy automobilů mýtnou branou



Krabicový graf (Box & Whiskers) Y



I. Variabilita

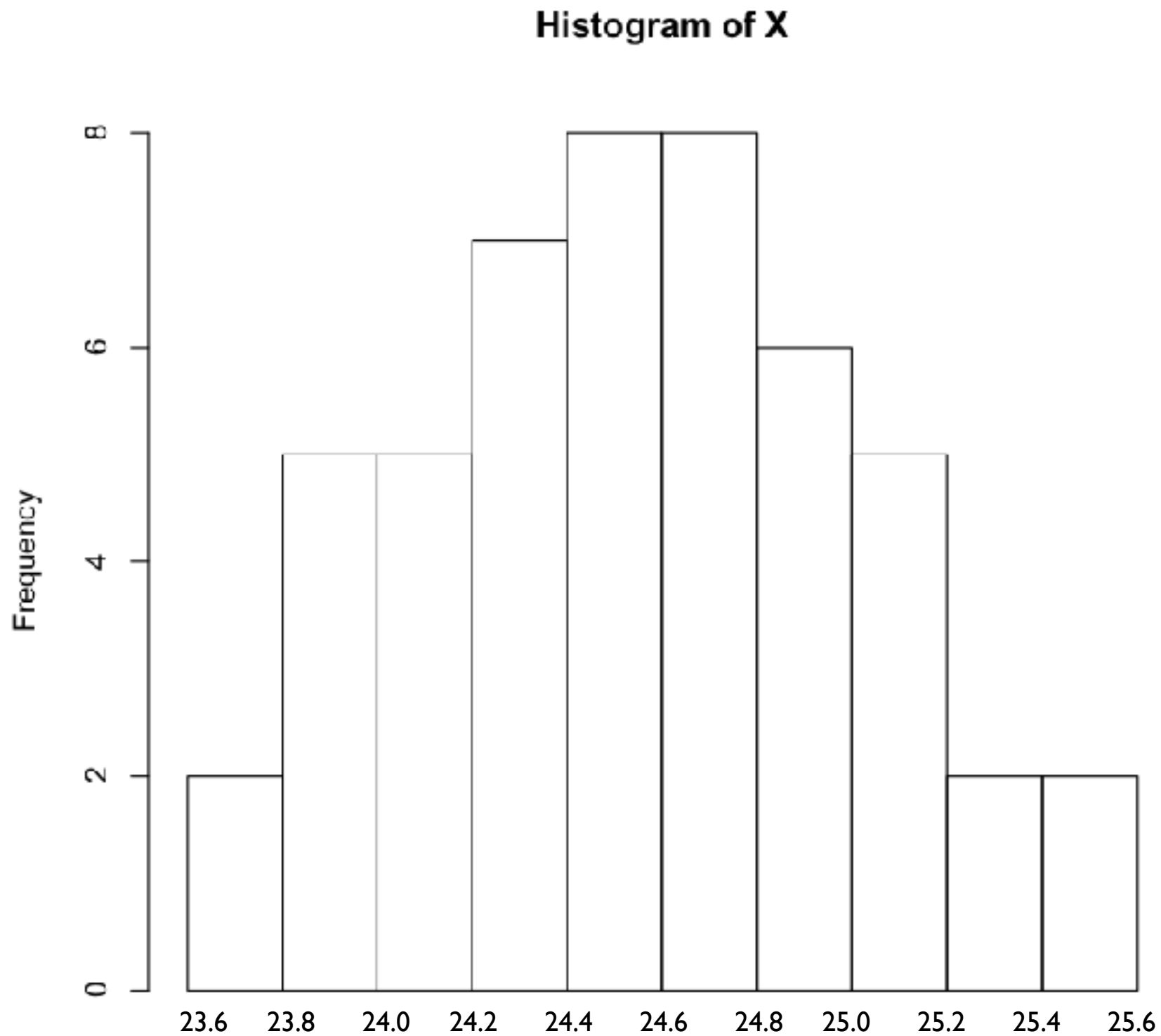
Balící automat na kávu

24.52586	24.17119	24.54486	24.44240	23.93455
24.20389	24.19974	24.34851	23.94024	24.21022
24.87474	25.06155	25.48924	25.32572	23.71721
24.61622	25.06676	24.90055	24.36213	24.98580
24.80591	24.20853	24.72623	24.64437	24.70405
23.97645	25.29837	24.46910	24.99453	25.42994
24.66147	24.75773	25.03970	24.44901	25.13285
24.40205	24.78721	23.83656	24.17186	23.65390
24.48244	24.68550	24.22988	23.83956	24.09777
24.52098	24.89240	24.25332	24.14259	25.12906



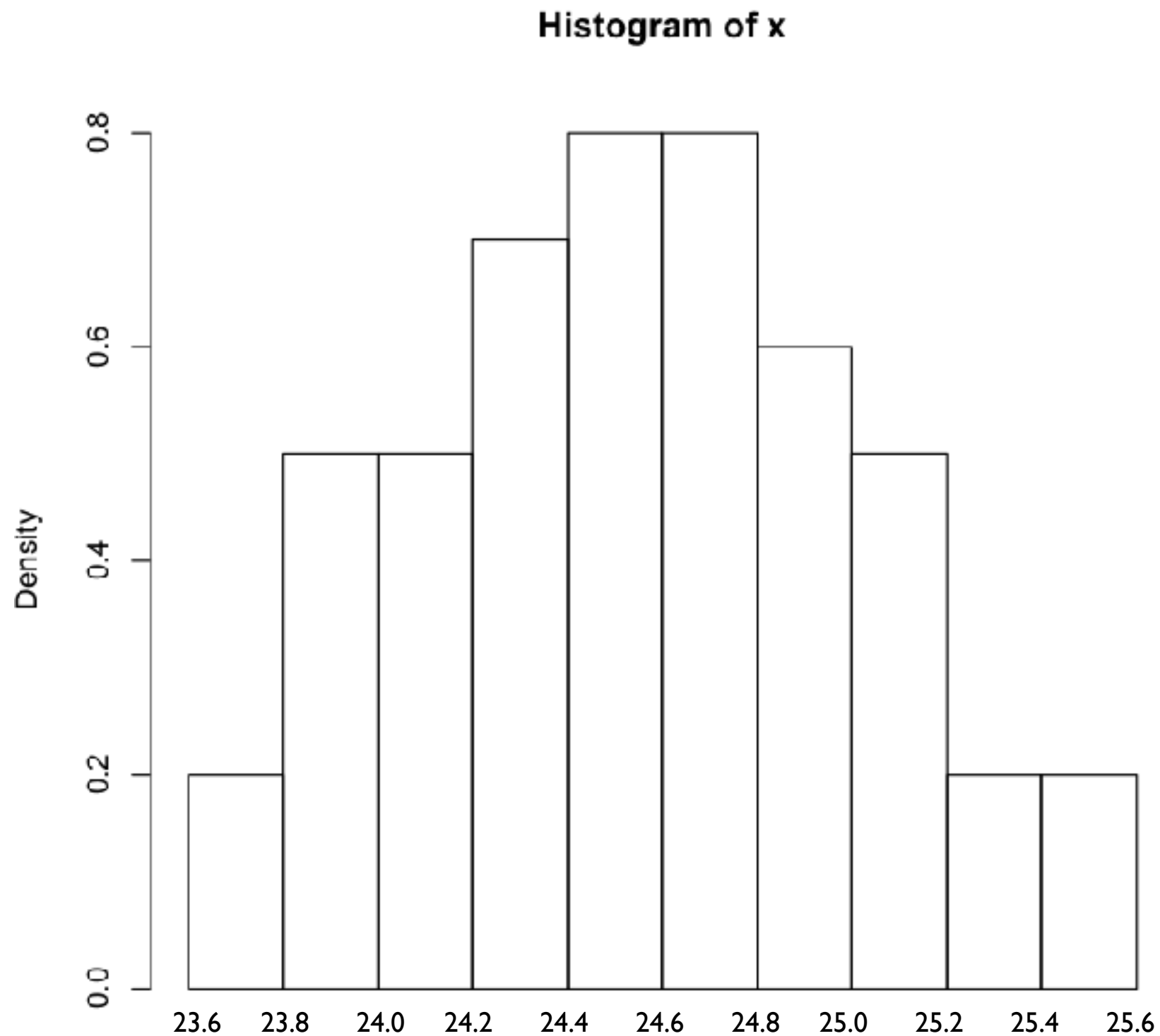
I. Variabilita

Balící automat na kávu



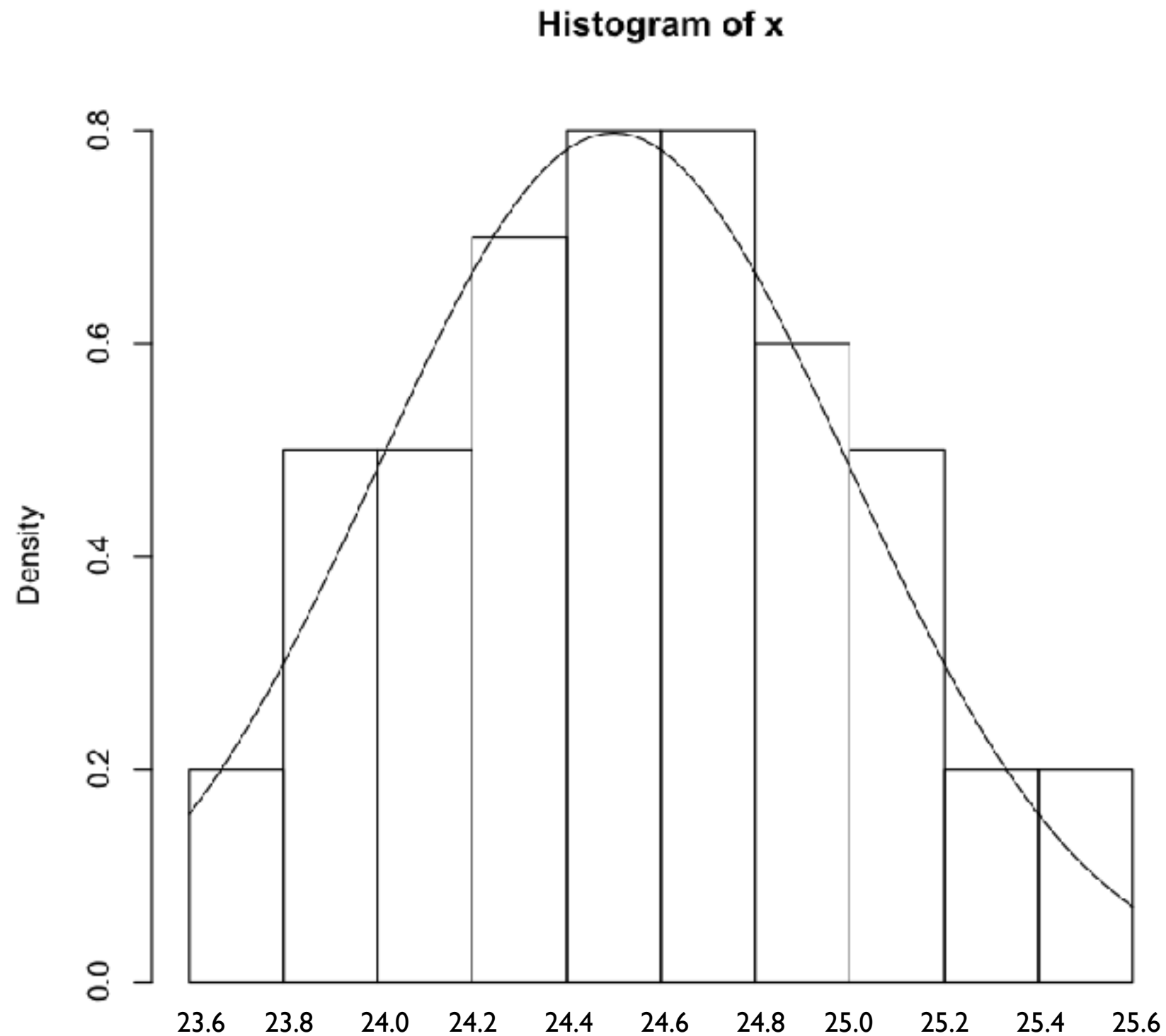
I. Variabilita

Balící automat na kávu



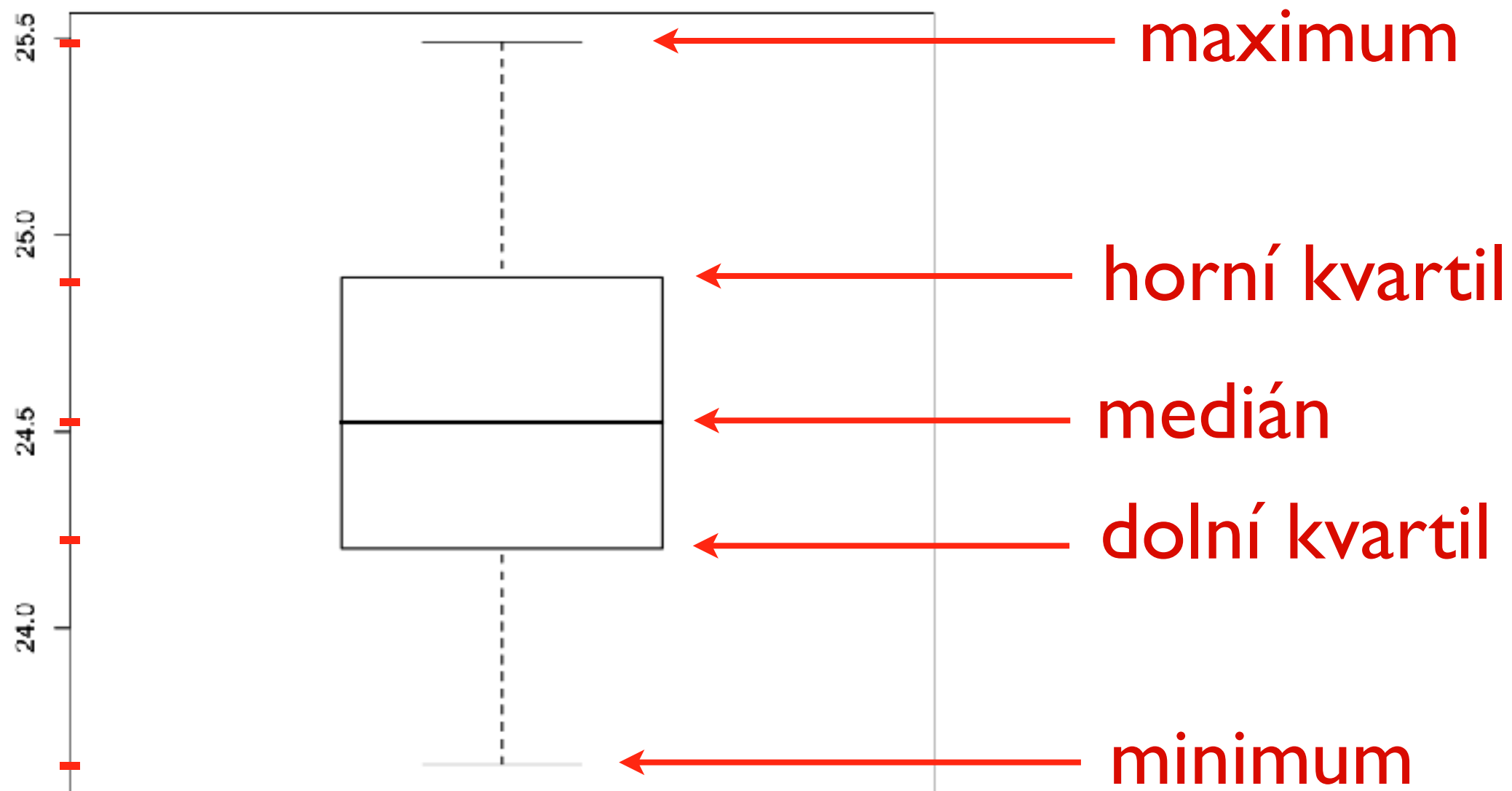
I. Variabilita

Balící automat na kávu



I. Variabilita

Balící automat na kávu



I. Variabilita

Kakuan (1100-1200)



Hledání býka



Nalezení stop



Spatření býka



Návrat domů



Zkrocení býka



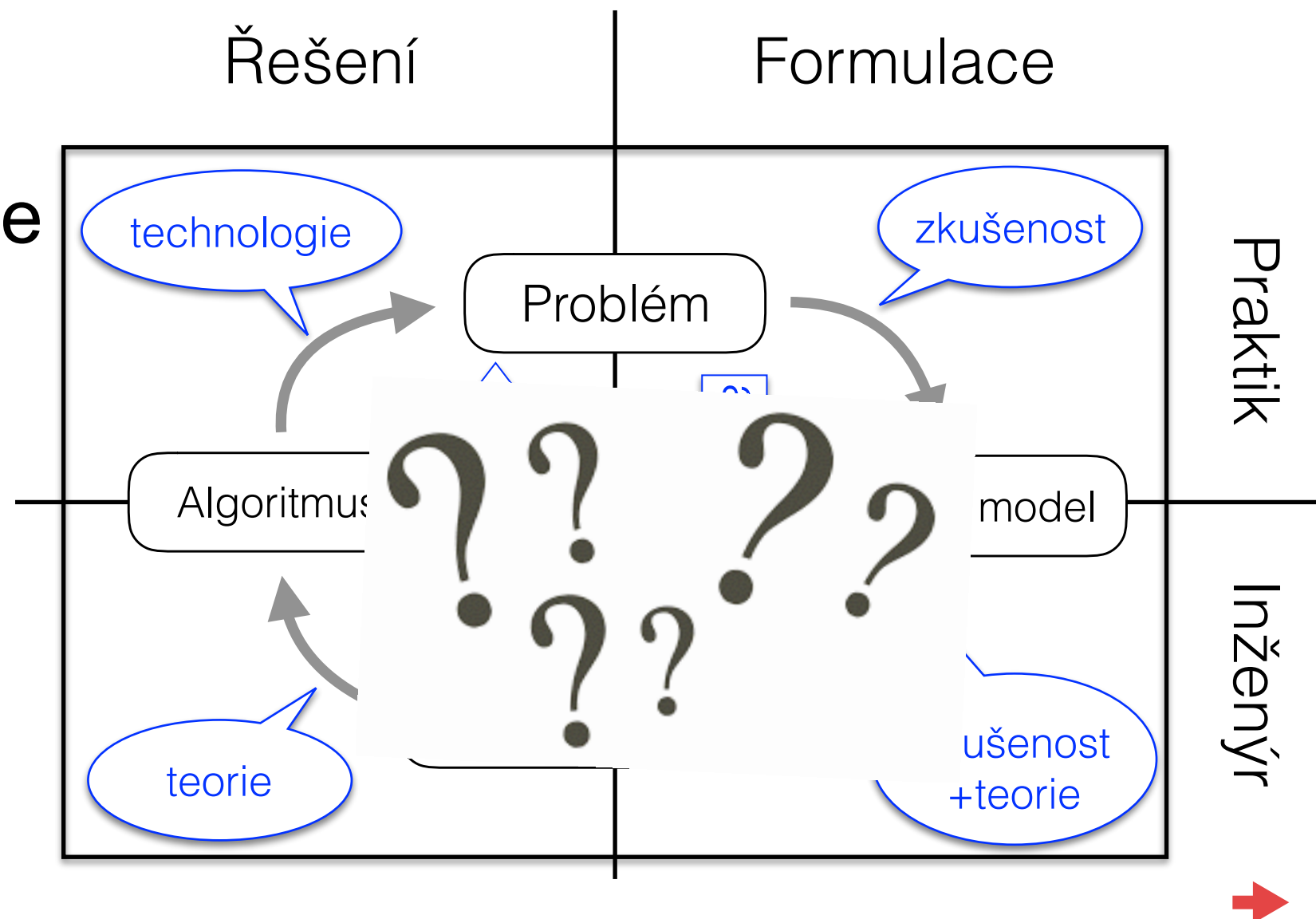
Chycení býka

<http://www.luzs.cz/deset-obrazu-kroceni-byka.html>



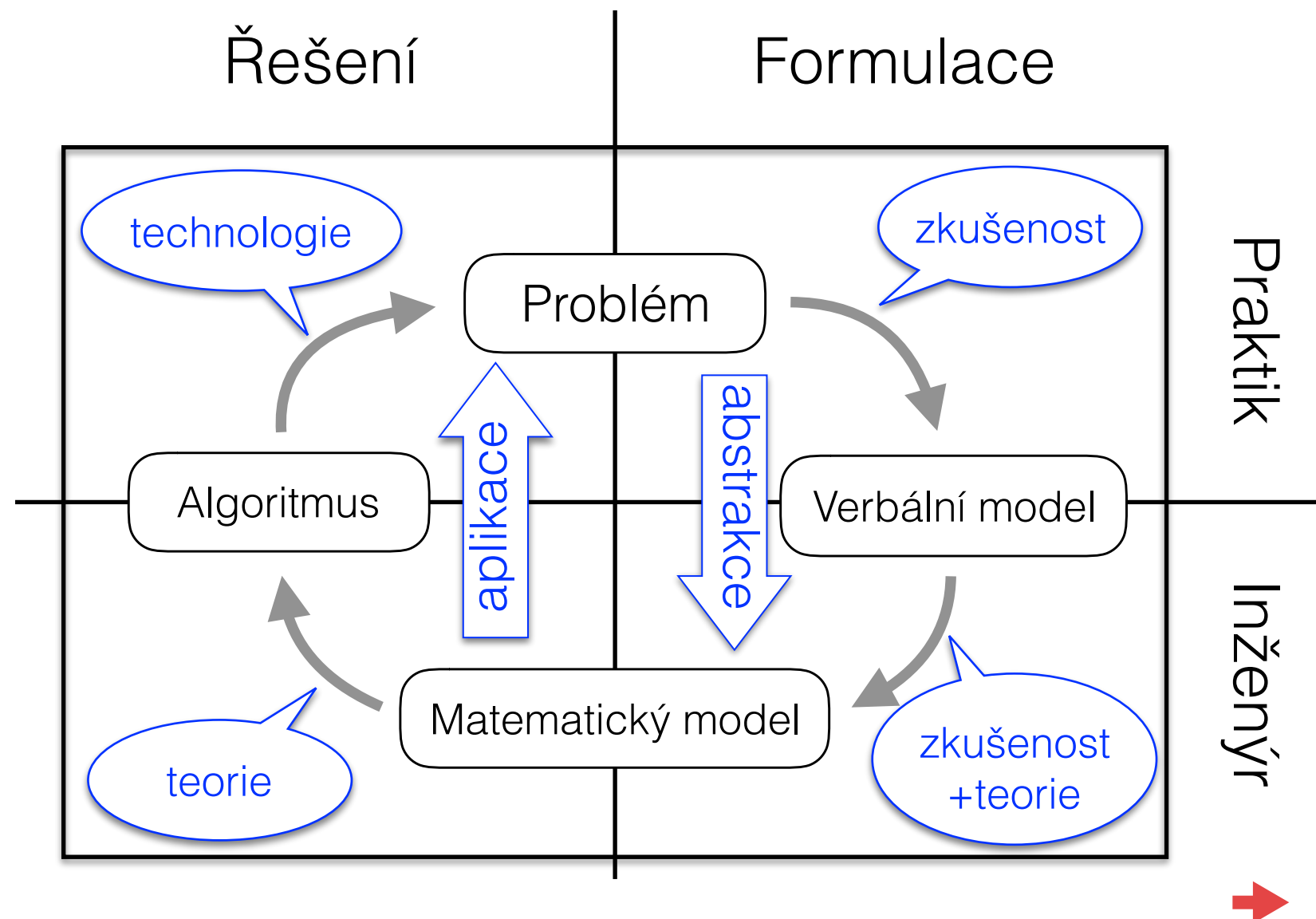
I. Úvod a motivace

- Variabilita
- Náhoda
- Model
- Experiment
- Statistická indukce



I. Úvod a motivace

- Variabilita
- **Náhoda**



I. Náhoda

<https://www.youtube.com/watch?v=uMTZMAIuQgg>

I. Náhodné jevy

- Náhodný experiment
vyústí v některý z různých, ale známých výsledků
- elementární náhodné jevy Ω
 Ω je množina konkrétních výsledků daného experimentu
- jevové pole $\mathcal{F}$
je to množina všech možných výsledků (podmnožin Ω)

$$(1) \quad \emptyset \in \mathcal{F}$$

$$(2) \quad A \in \mathcal{F} \Rightarrow \bar{A} \in \mathcal{F}$$

$$(3) \quad A_1, A_2, \dots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$$

- $(\Omega, \mathcal{F})$ náhodný prostor
charakterizuje náhodný experiment

Náhodné x deterministické jevy



I. Náhodné jevy

https://www.youtube.com/watch?v=VWNm5cmSB1hY&list=RDVWNm5cmSB1hY&start_radio=1&t=0



I. Kvantifikace náhody

Na počátku bylo hledání „matematické naděje“



Antoine Gombaud, Chevalier de Méré
(1607 – 1684)



I. Kvantifikace náhody

Na počátku bylo hledání „matematické naděje“



Antoine Gombaud, Chevalier de Méré
(1607 – 1684)

První hra:

Spočívala v hodu jednou kostkou. Chevalier de Méré přijímal sázky na to, že hodí minimálně jednu šestku ve čtyřech po sobě následujících hodech. Věděl, že pravděpodobnost padnutí šestky je v každém hodu $1/6$. Domníval se, že jeho šance na padnutí šestky ve čtyřech hodech je tedy $(1/6) \times 4 = 2/3$.

Druhá hra:

Druhá hra spočívala v hodu dvěma kostkami. Chevaliera de Méré byl úspěšný, pokud se mu alespoň jedenkrát podařilo hodit 2 šestky ve 24 hodech. Opět věděl, že pravděpodobnost vrhnutí dvou šestek v jednom hodu je $1/36$ (pouze jedna možnost z 6×6 možných případů). Předpokládal, že jeho šance je tedy $(1/36) \times 24$ a tudíž opět $2/3$.



Přesto při druhé hře v kostky utrpěl Chevalier de Méré značné finanční ztráty. Jsou šance na výhru v obou kostkových hrách skutečně stejné? V čem byl jeho předpoklad chybný?

I. První formulace teorie pravděpodobnosti



Blaise Pascal (1623 - 1662)

Rozdělení sázky:

Dva hráči hrají sérii lichého počtu N her a vyhraje ten, který dosáhne vícekrát vítězství. Na vítěze je vyhlášena sázka. Hra je však přerušena za stavu $m:n$ a je třeba sázku rozdělit mezi oba hráče.

Jaký je nejspravedlivější poměr rozdělení sázky?



Pierre de Fermat (1601 - 1665)

Úloha o rozdělení sázky je jednou z nejstarších dokumentovaných úloh pravděpodobnosti a je připisována Blaise Pascalovi a Piere de Fermatovi z roku 1654. Ve skutečnosti je však mnohem starší a jsou známy její verze z konce XV. století. [[Informační bulletin ČStS 2009 \(2\)](#)]

Dva hráči hrají sérii 9 her a vyhraje ten, kdo první dosáhne 5 vítězství. Na vítěze je vyhlášena sázka. Hra je však přerušena za stavu 4:2 a je třeba spravedlivě rozdělit sázku mezi oba hráče. Jak?



I. Pravděpodobnostní úlohy

Úloha o ruinování hráče

Dva hráči A a B hrají sérii her. Při každé hře hráči vsadí jednu korunu. Hráč A vyhraje s pravděpodobností p a prohraje s pravděpodobností $1-p$. Na počátku má hráč celkem x korun, hráč B má y korun. Hry jsou vzájemně nezávislé. Hra se hraje tak dlouho, dokud mají oba co vsadit. Tedy hra končí v okamžiku, kdy je jeden z hráčů zruinován. Jakou mají hráči A a B pravděpodobnost zruinování?

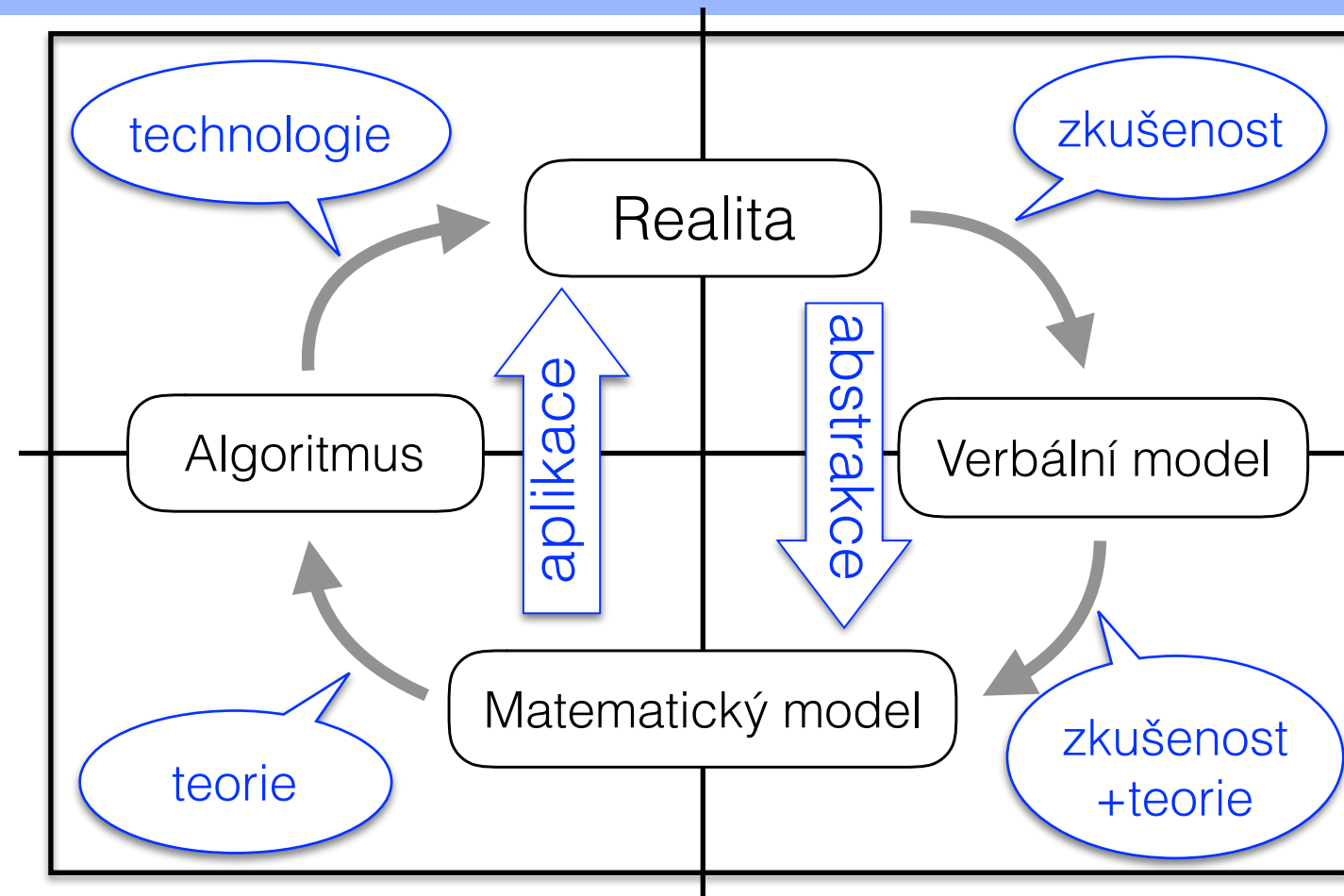
Další z nejznámějších pravděpodobnostních úloh, kterou lze nalézt například v učebnici teorie pravděpodobnosti W. Feller z roku 1957.

<http://www.statapol.cz/oldstat/bulletiny/ib-93-4.pdf>



I. Úvod a motivace

- Variabilita
- Náhoda
- **Model**
 - je zjednodušením reality (zanedbáním „šumu“)
 - zaměřený na problém (měl by vystihovat hlavní podstatu problému)
 - pomáhá porozumět složitější realitě
 - nikdy není přesný



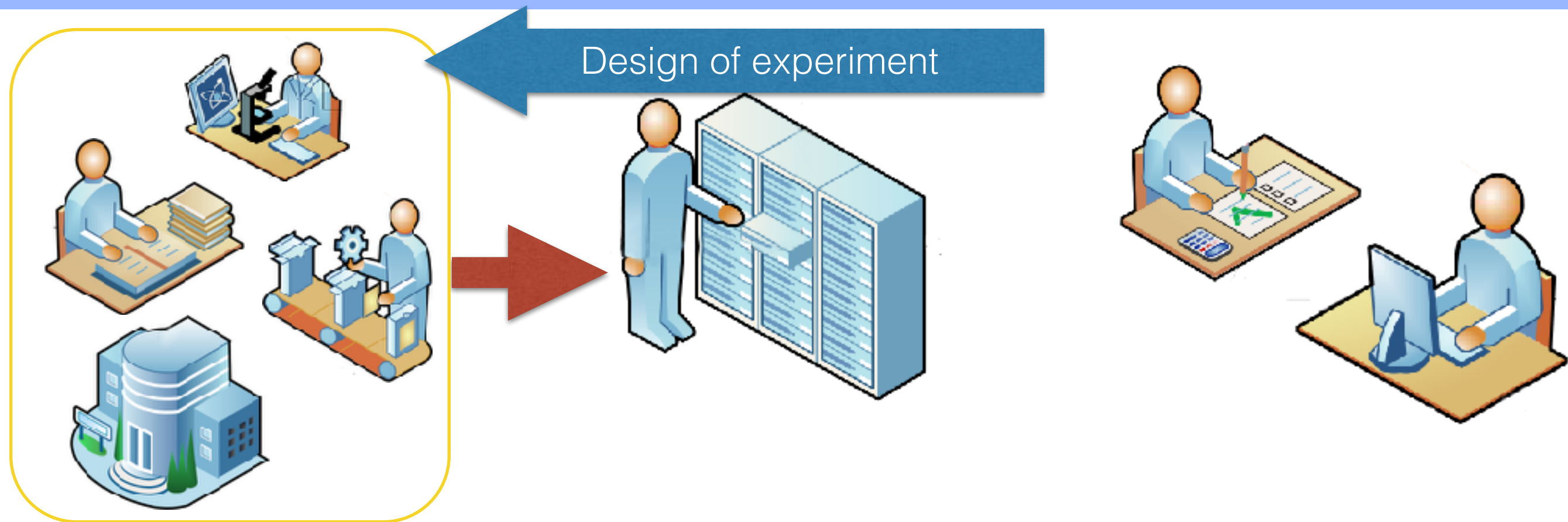
I. Model

Při výběru a konstrukci modelu musíme respektovat některé vlastnosti:

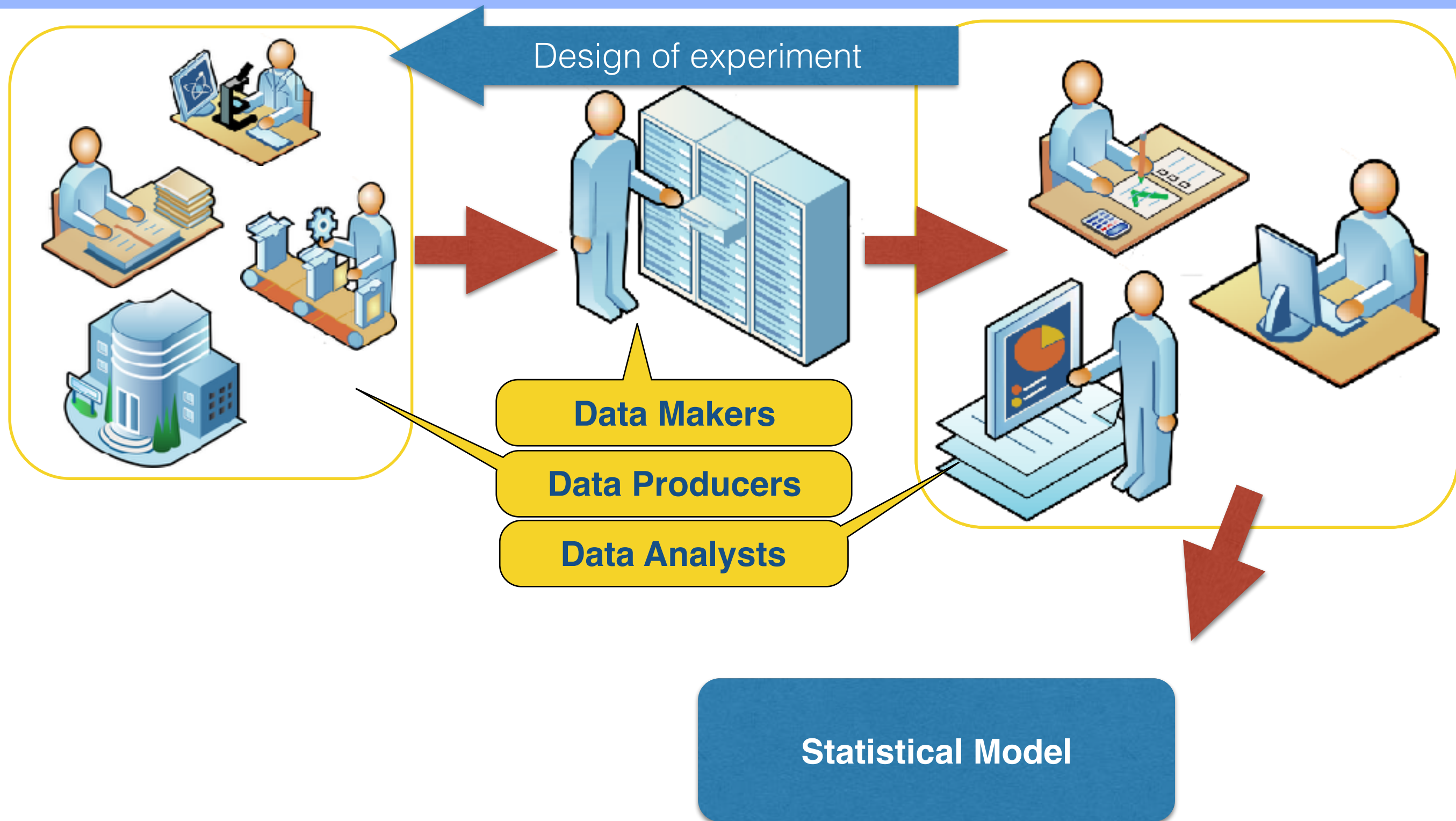
- **falšovatelnost (falsifiability)**: zda existují potenciální pozorování, která jsou neslučitelná s modelem;
- **přiměřenost vysvětlení (explanatory adequacy)**: zda teoretický výklad modelu pomáhá pochopit nejen pozorovaná data, ale také dosavadní zjištění;
- **interpretovatelnost (interpretability)**: zda jsou komponenty modelu, zejména jeho parametry, srozumitelné a jsou v souladu se známými procesy;
- **věrnost (faithfulness)**: zda schopnost modelu zachytit základní zákonitosti pochází z teoretických principů, které se model snaží implementovat a nikoli z náhodných rozhodnutí provedených v průběhu jeho tvorby;
- **dobrá shoda (goodness of fit)**: zda model dostatečně dobře vyhovuje pozorovaným datům;
- **zobecnitelnost (generalizability)**: zda model poskytuje dobrou předpověď budoucích pozorování;
- **složitost (complexity or simplicity)**: zda jsou sledovaná data modelem popsána co možná nejjednodušším způsobem;



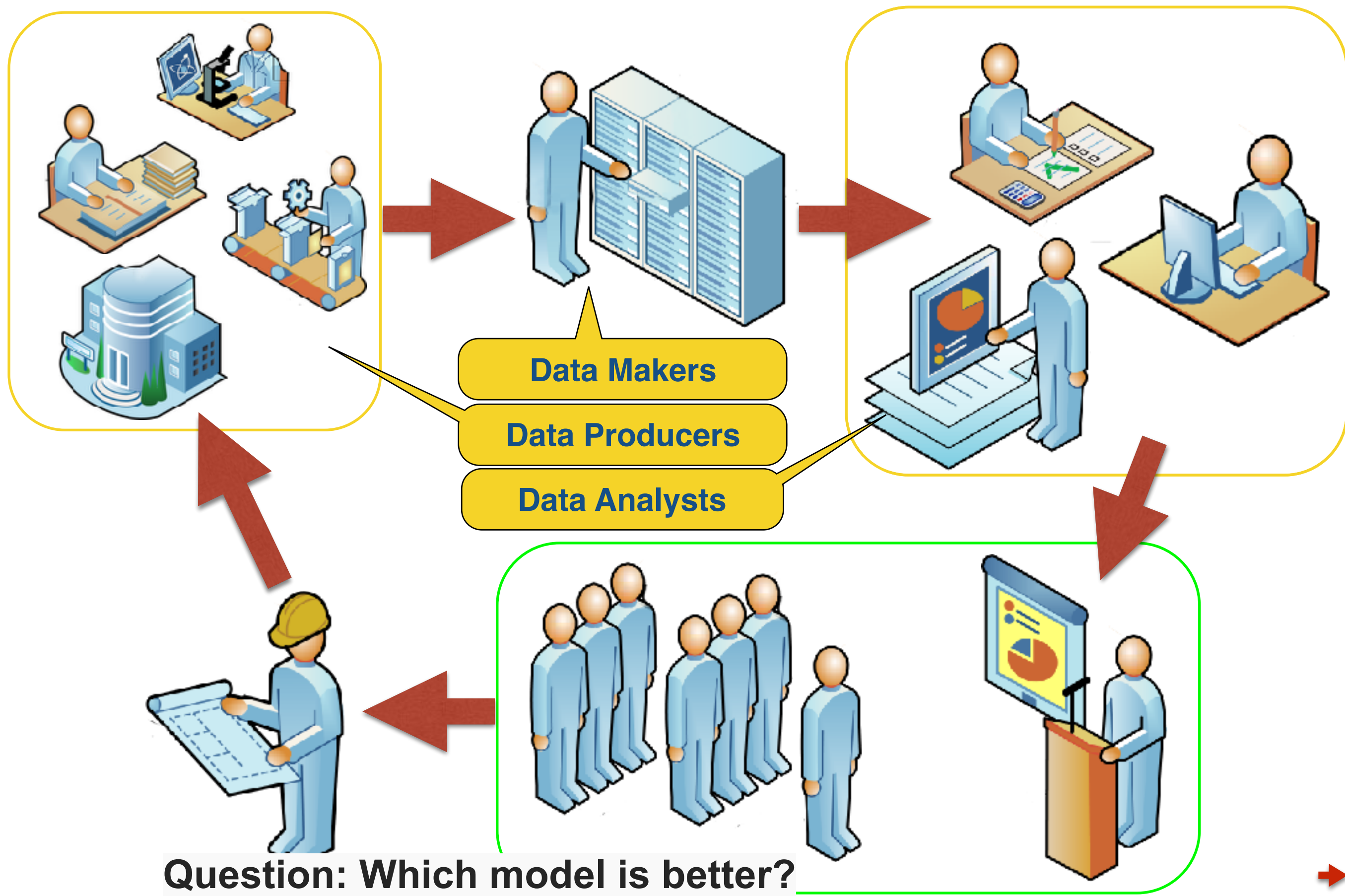
I. Životní cyklus dat



I. Životní cyklus dat



I. Životní cyklus dat



I. Model

... všechny modely jsou přibližné. V podstatě všechny modely jsou špatné, ale některé z nich jsou užitečné. Nicméně, přibližný charakter modelu musíme mít vždy na paměti, ...



George E. P. Box (1919 - 2013)



Norbert Wiener (1894 – 1964)

Nejlepším modelem kočky je kočka, nejlépe ta samá.



I. Ocam's razor

Máte-li dva konkurenční modely, které vám dávají přesně stejné předpovědi, potom lepší je ten jednodušší.

Bohužel v mnoha oblastech se často zapomíná, že modely nemusejí nutně být přesnou kopií reality, aby umožnily nové objevy a předpovědi! Neztrácejte čas vytvářením komplikovaného modelu, který vyžaduje přesná měření velmi mnoha proměnných. Skutečný génius přináší jednoduchý model, který postačí k řešení problému.



Isaac Newton (1642 – 1727)



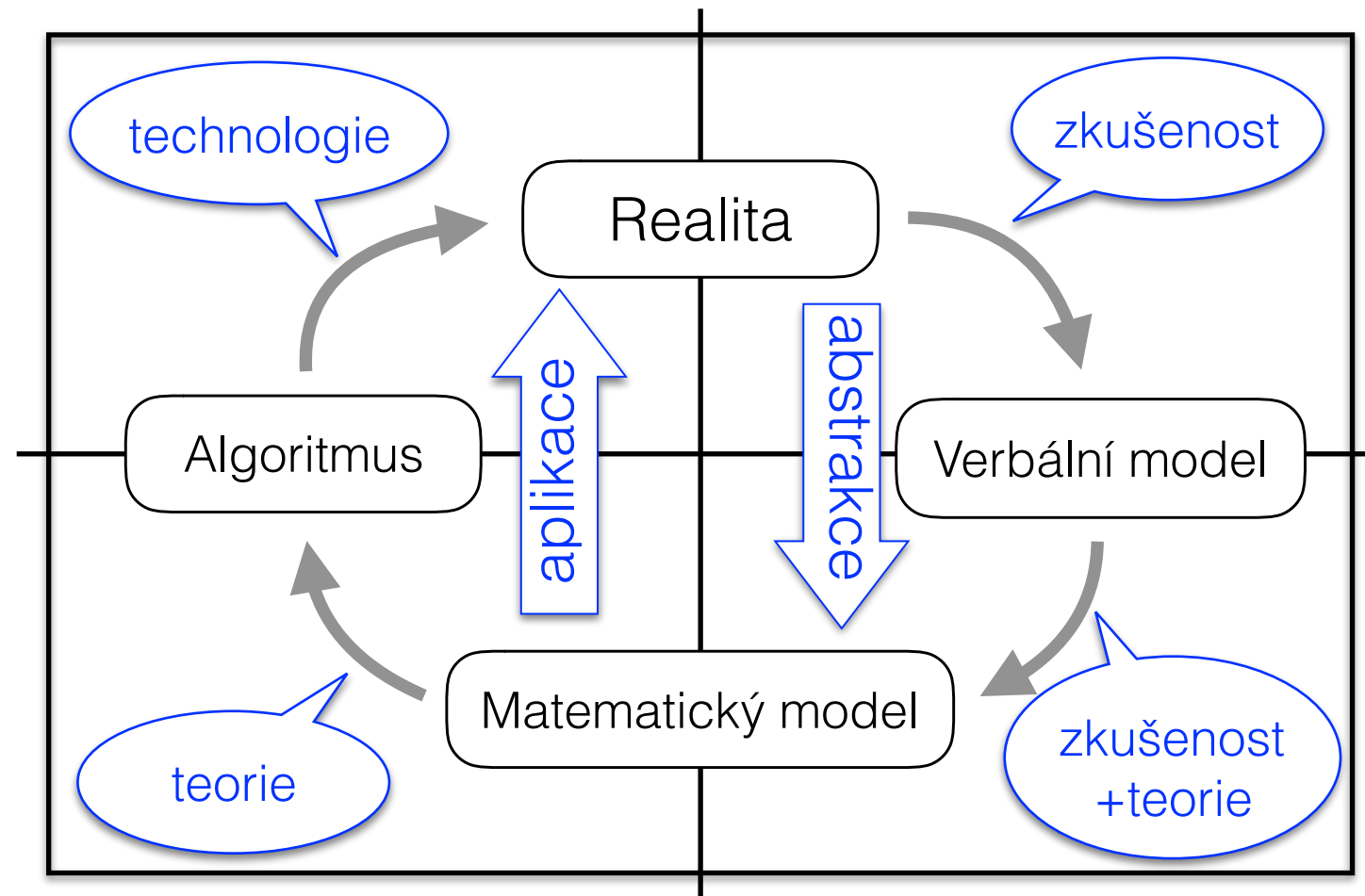
William Ockham (1287 – 1347)

Nepřipouštějme více příčin přírodních jevů, než ty, které jsou pravdivé a současně dostatečné k tomu, aby vysvětlily výskyt a vlastnosti těchto jevů.



I. Úvod a motivace

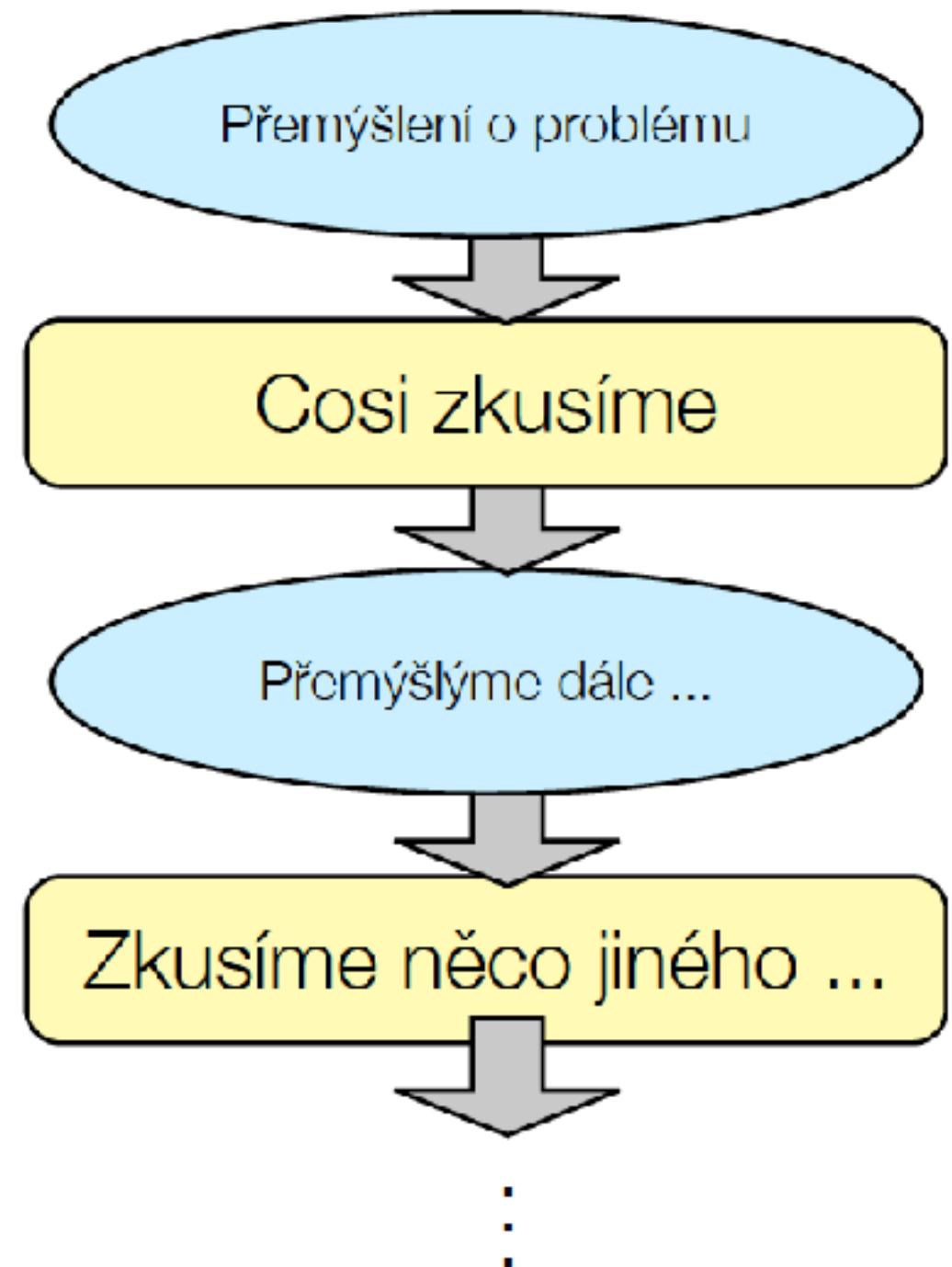
- Variabilita
- Náhoda
- Model
- **Experiment**



I. Experiment

Nesystematický přístup k řešení problémů - metoda “pokusu a omylu”:

- funguje v jednotlivých případech a jen s několika málo lidmi
- často čekáme velmi dlouho na výsledek
- vyžaduje zkušenost a intuici
- lze jej využít pouze v omezeném okruhu aplikací

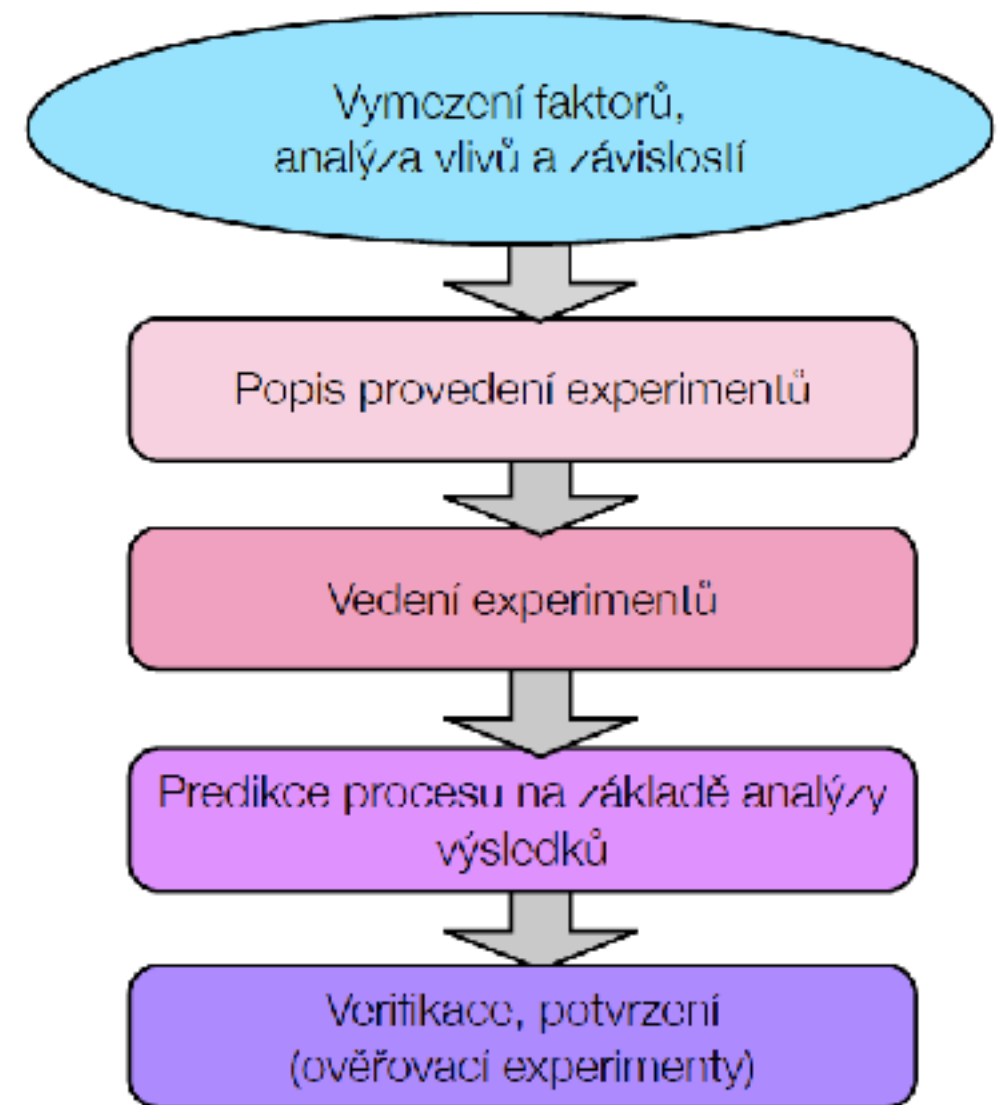


I. Experiment

Systematický přístup k řešení problémů - plánování experimentu:

Experiment je test nebo série testů (pokusů), zpravidla provedená za účelem zvýšení kvality produktu nebo procesu, případně zvýšení jejich efektivity.

- stanovení charakteristik procesu a jeho optimalizace
- vyhodnocení vlastností materiálů
- návrh a vývoj produktů
- stanovení tolerance komponent a vstupních veličin

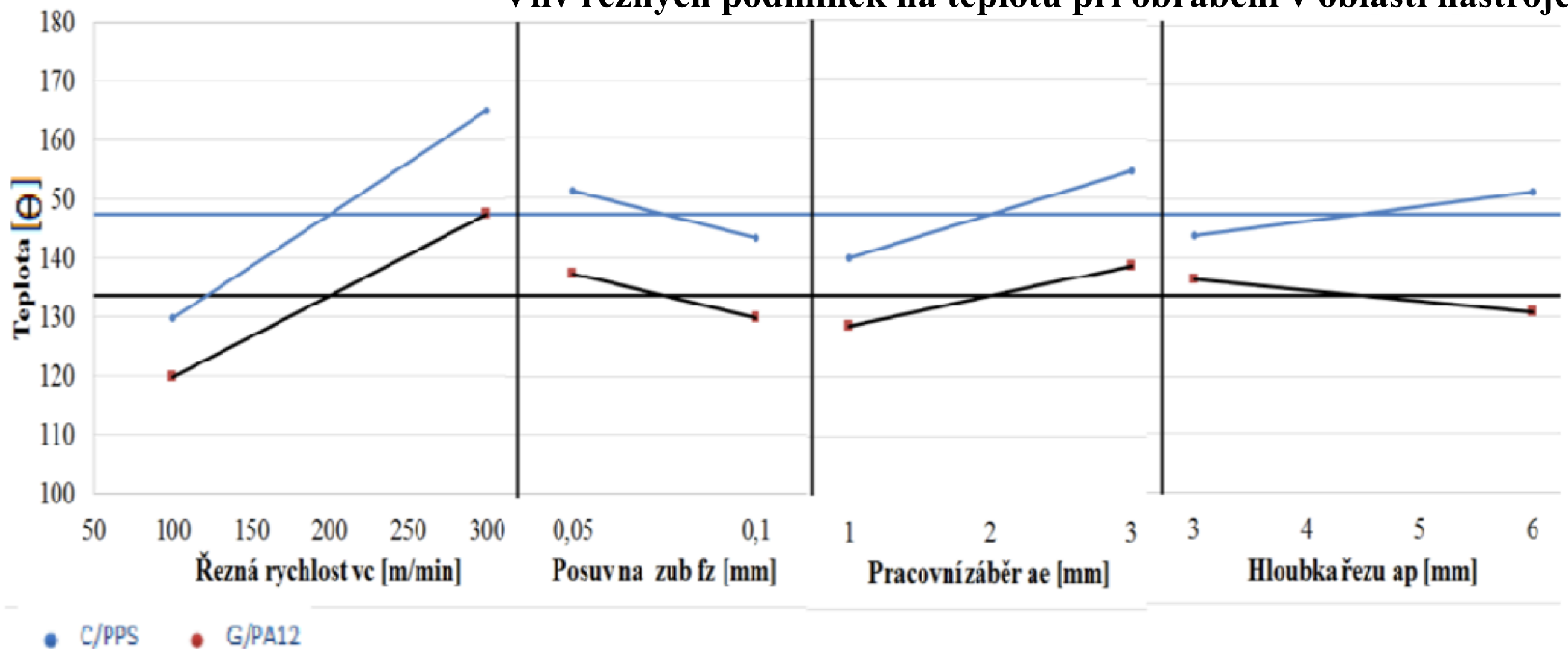


I. Experiment

Vliv řezných podmínek na teplotu při obrábění

- řezná rychlost,
- velikost posuvu na zub,
- pracovní záběr,
- hloubka řezu

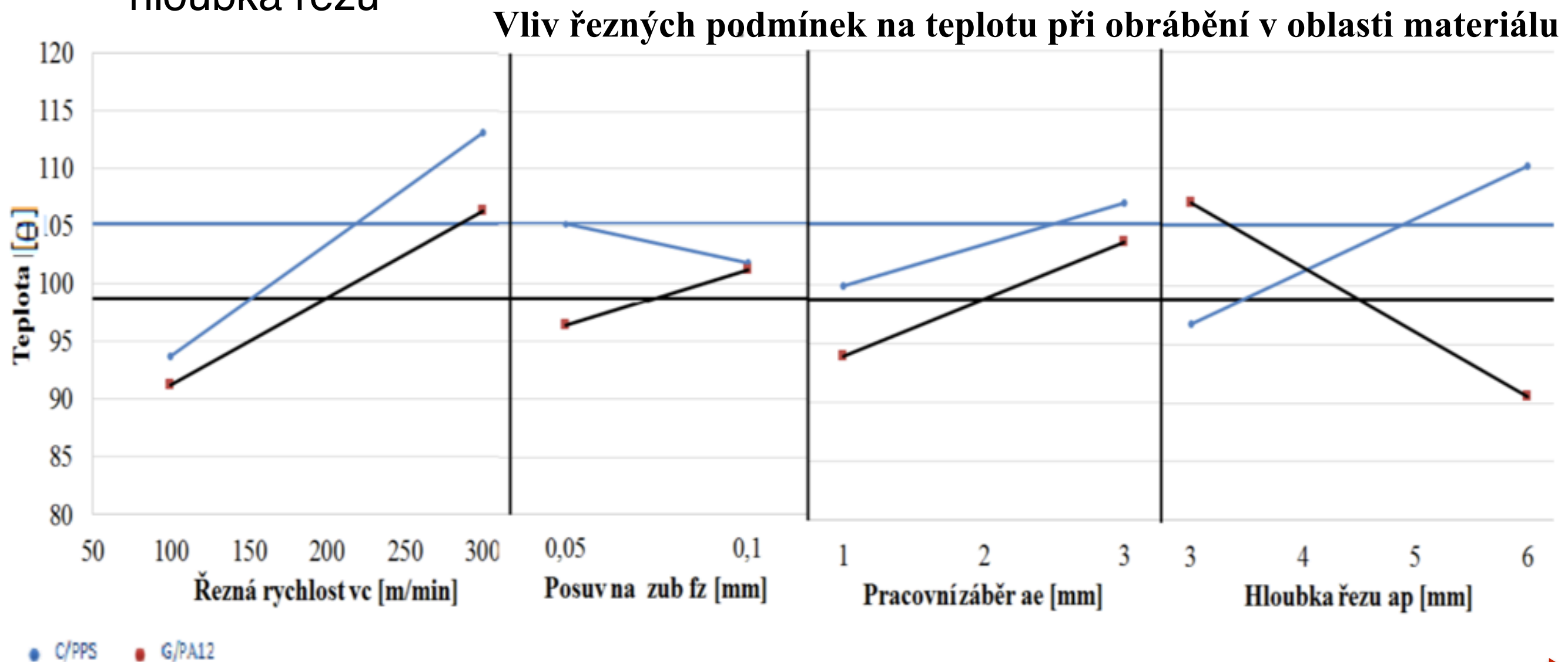
Vliv řezných podmínek na teplotu při obrábění v oblasti nástroje



I. Experiment

Vliv řezných podmínek na teplotu při obrábění

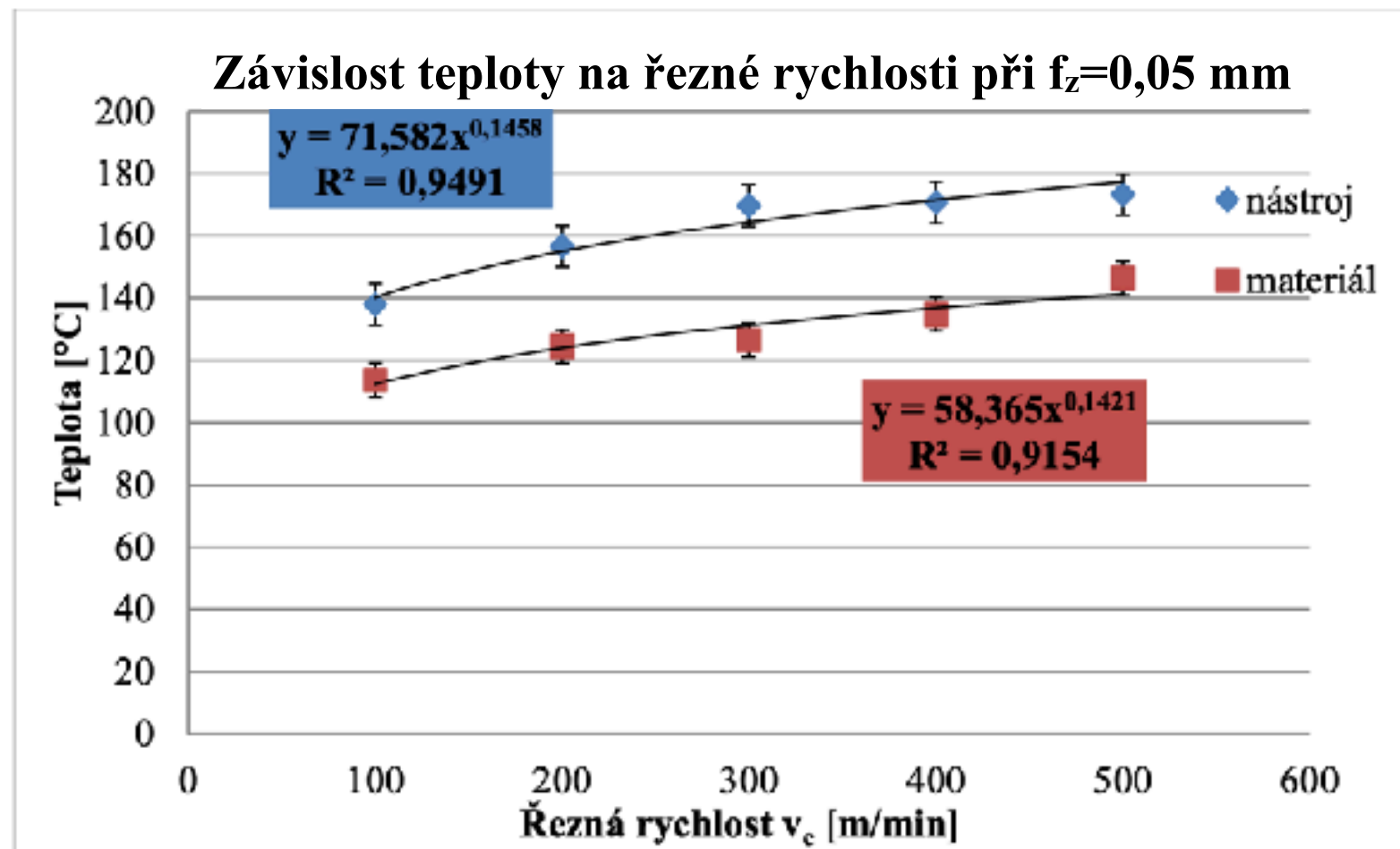
- řezná rychlost,
- velikost posuvu na zub,
- pracovní záběr,
- hloubka řezu



I. Experiment

Vliv řezné rychlosti na teplotu v místě řezu

$$T = C \cdot v_c^z + \varepsilon$$



I. Statistické vyhodnocení experimentu

Konzultovat se statistikem poté, co je experiment dokončen, je často totéž, jako požádat lékaře, aby provedl pitvu. Možná bude moci sdělit, nač experiment zemřel.



Ronald Fischer (1890 - 1962)

Naší povinností je provádět shrnutí a jasně a srozumitelně formulovat své závěry, abychom umožnili dalším svobodně a nezávisle je využívat při vytváření vlastní rozhodnutí.

Induktivní odvozování je jediný známý proces, při kterém přicházejí na svět nové důležité poznatky ➡

I. Pár citátů na závěr první části



Walter Edward Deming (1900 - 1993)

Pokud nemůžete popsat to, co děláte jako proces,
pak nevíte, co děláte.

Nestačí dělat věci co nejlépe; musíte vědět,
co máte dělat, a pak to dělat co nejlépe.

Ovlivňujme příčiny, nikoli důsledky.

Věříme pouze v Boha, všechno ostatní musí přinést data.



I. Úvod a motivace

<https://www.ceskatelevize.cz/ivysilani/872236-hrisni-lide-mesta-prazskeho/267531305460006-pripad-licheho-strevice/titulky#t=15m28s>