EVIDENCE-BASED MANAGEMENT: HR ANALYTICS CASE STUDY

Michal Jirásek

Podniková informatika pro manažery

PROGRAM DNEŠNÍHO SETKÁNÍ

1. Kahoot
2. Úvod do Evidence-based management a zadání případové studie
3. Rešerše – online a offline
4. Power BI a další přehledové statistiky
5. Formulace modelů
6. SPSS Modeler a interpretace výsledků
7. Napočítání modelů a vyhodnocení

PROGRAM DNEŠNÍHO SETKÁNÍ

Cíle dnešního setkání

Zkušenosti s analýzou dat vs „španělská vesnice“?

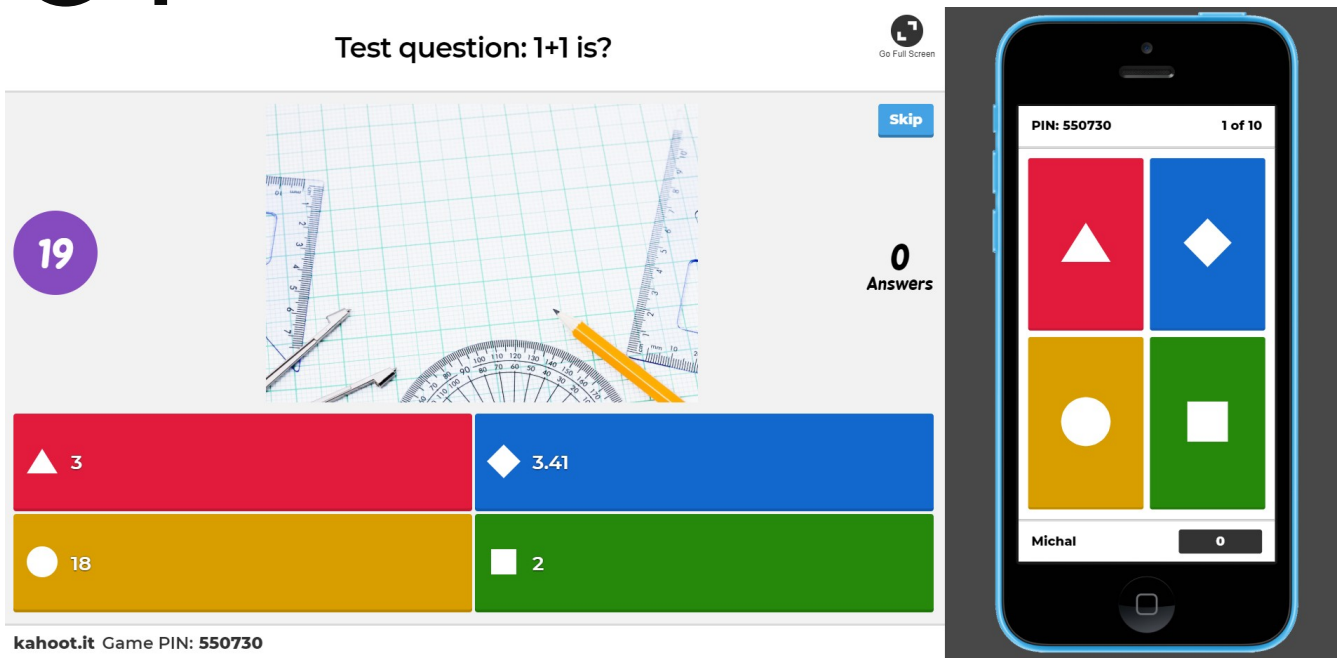
Zkušenosti z/s HR?

Nainstalovaný software?

Game PIN

Enter

KAHOOT



kahoot.it

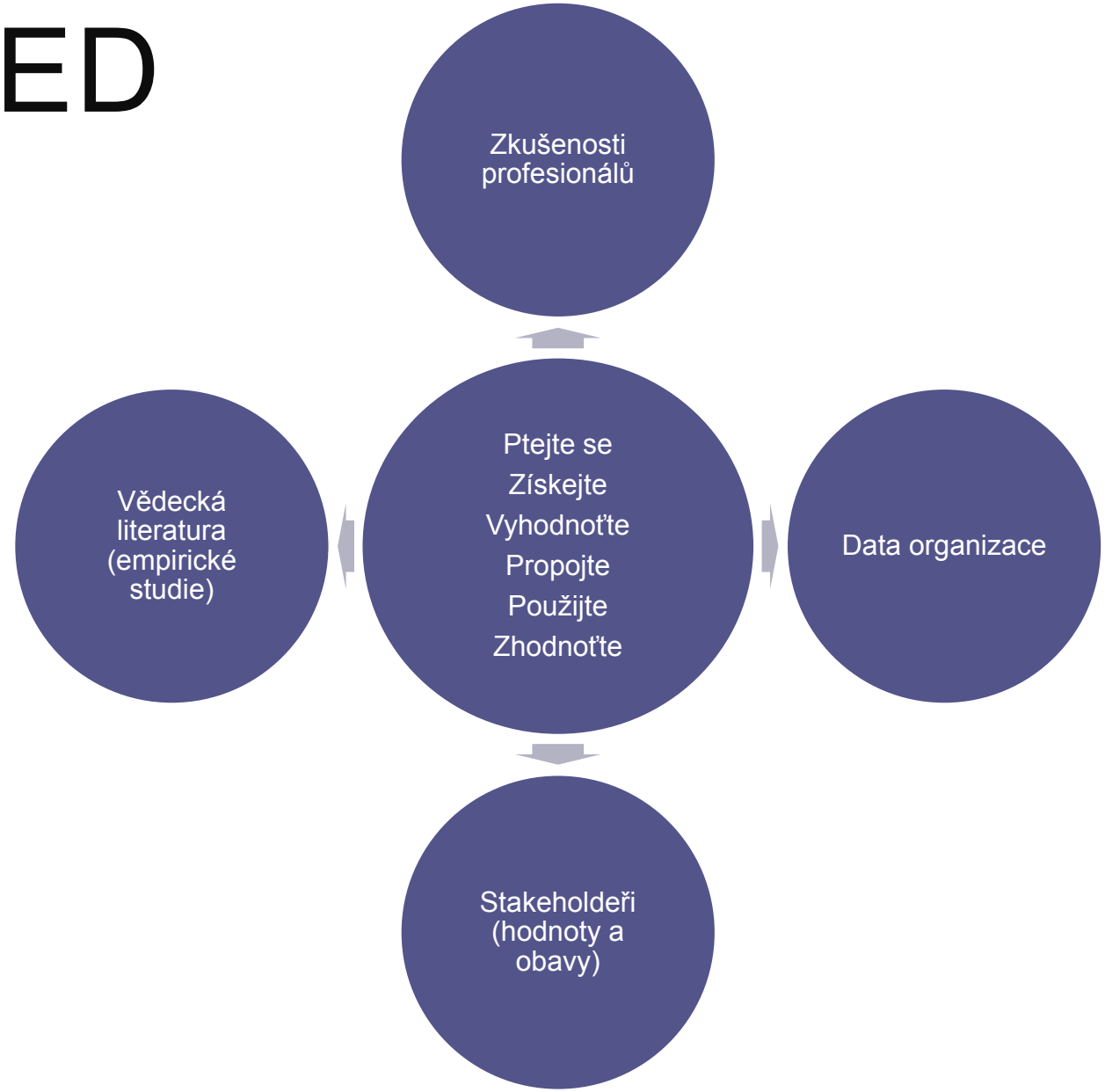
Pravda/Lež + Multiple-choice otázky (stačí jedna správně)

Časový limit

EVIDENCE-BASED MANAGEMENT

Motivace: Lidské rozhodování je zkreslené (v podnicích o nic méně)

> Rozhodování by mělo být založeno na těch nejlepších dostupných zdrojích



PŘÍPADOVÁ STUDIE

Jaké faktory ovlivňují odchody zaměstnanců z podniku?

Data: Kaggle.com

Podklady vloženy ve Studijních materiálech:

- HR Analytics Case Study.pdf
- Data.zip
  - data_dictionnary.xlsx – popis proměnných

REŠERŠE

Jak hledáte odborné informace?

- S jakými klíčovými slovy budete pracovat?

S jakými zdroji pracujete?

Zkuste rychle najít odpovědi na otázky (20-25 minut)

VIZUALIZACE DAT

Proč je důležité data vizualizovat?

Vhodné nástroje pro vizualizaci dat?

Ukázka Power BI

POPISNÉ STATISTIKY A KORELACE

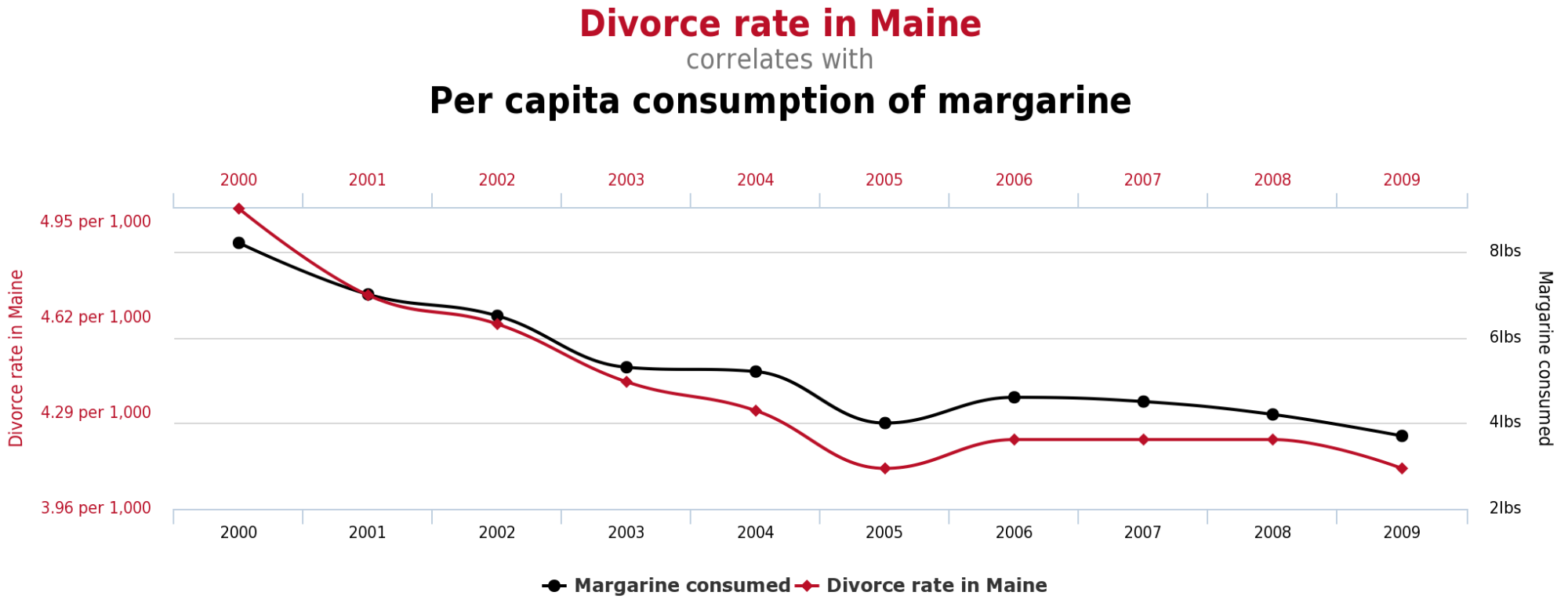
Co nám říkají?

Podstata regrese (OLS)?

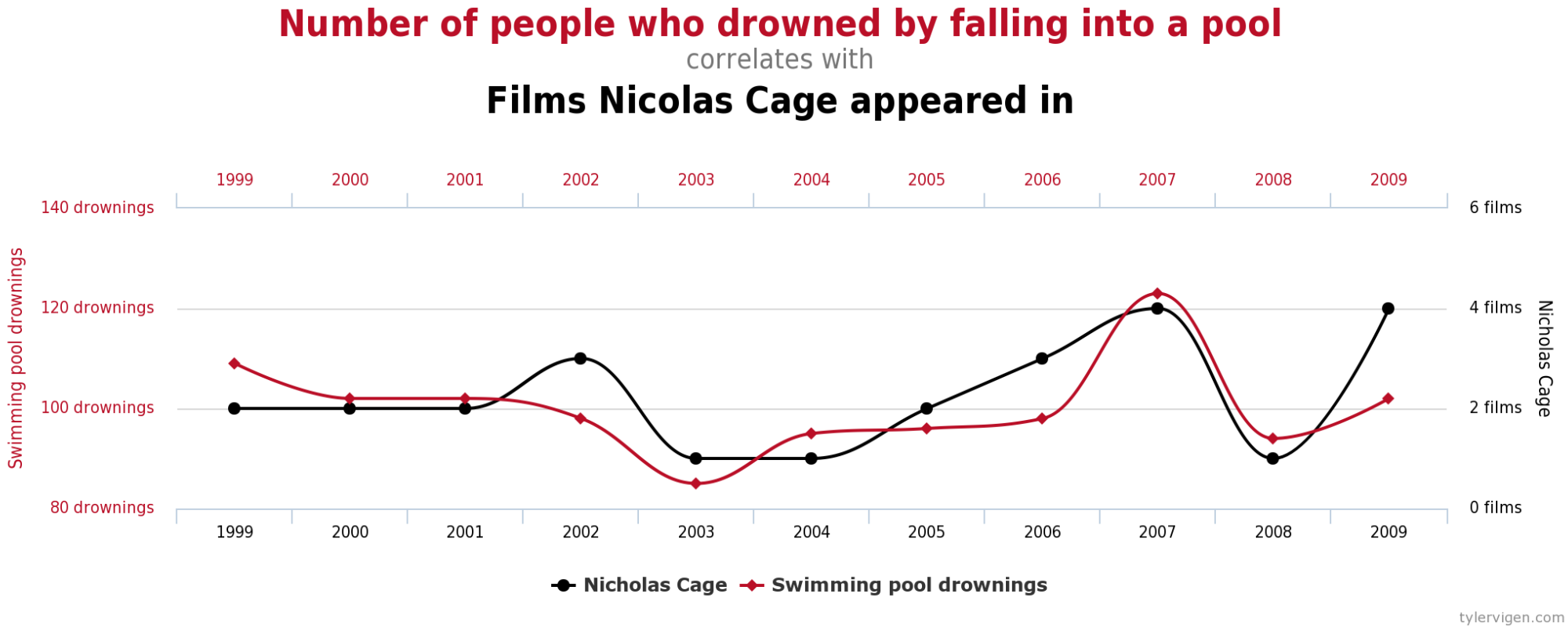
Jaký je jejich vztah k předpokladům a vlastnostem regrese (multikolinearita, normalita, vlivná pozorování)

Asociace (falešná) vs kauzalita

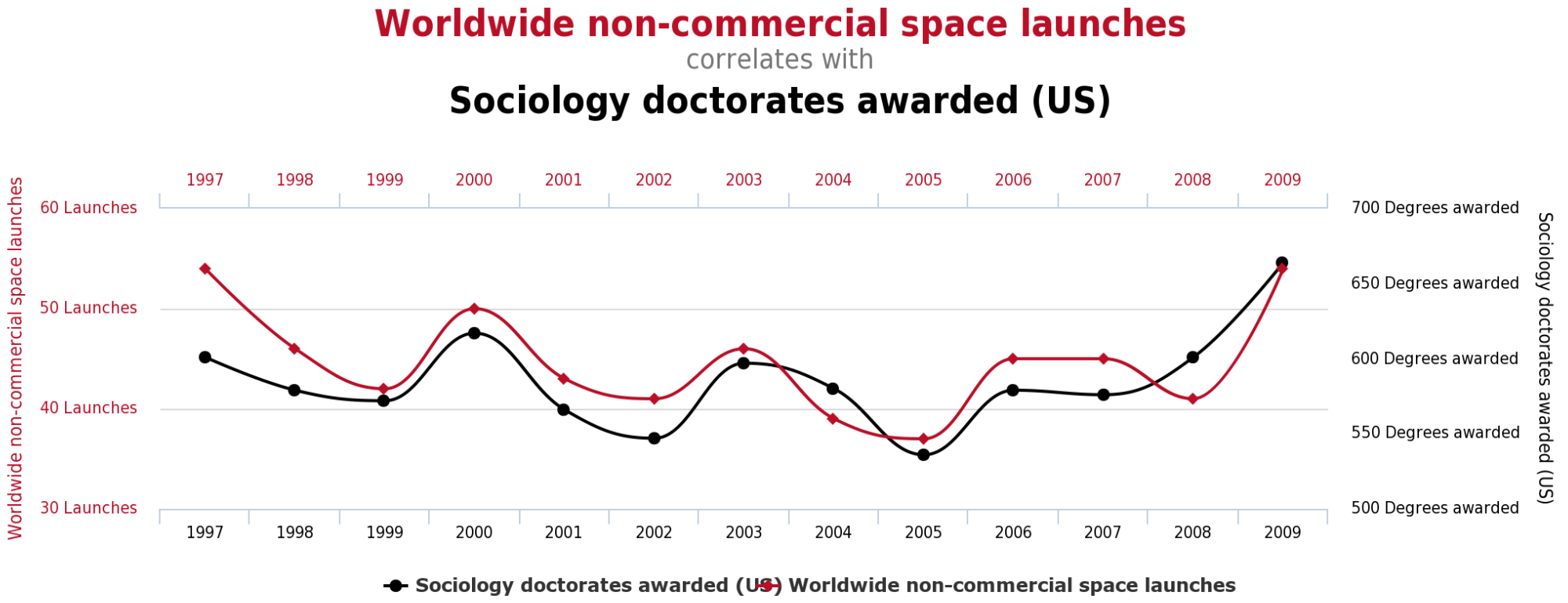
FALEŠNÉ KORELACE (SPURIOUS CORRELATIONS)



FALEŠNÉ KORELACE (SPURIOUS CORRELATIONS)



FALEŠNÉ KORELACE (SPURIOUS CORRELATIONS)



	N	Minimum	Maximum	Mean	Std. Deviation
Age	4410	18	60	36,92	9,133
DistanceFromHome	4410	1	29	9,19	8,105
Education	4410	1	5	2,91	1,024
JobLevel	4410	1	5	2,06	1,107
MonthlyIncome	4410	10090	199990	65029,31	47068,889
NumCompaniesWorked	4391	0	9	2,69	2,499
PercentSalaryHike	4410	11	25	15,21	3,659
StockOptionLevel	4410	0	3	,79	,852
TrainingTimesLastYear	4410	0	6	2,80	1,289
YearsAtCompany	4410	0	40	7,01	6,125
YearsSinceLastPromotion	4410	0	15	2,19	3,222
YearsWithCurrManager	4410	0	17	4,12	3,567
WorkLifeBalance	4372	1	4	2,76	,706
JobInvolvement	4410	1	4	2,73	,711
PerformanceRating	4410	3	4	3,15	,361
Valid N (listwise)	4353				

FORMULACE MODELŮ

Proč se formulují modely předem?

Co je to hypotéza?

Proč nezahrnout všechny proměnné?

„Overfitting“ (trénovací vs testovací množina)

FORMULACE MODELŮ

Formulujte v týmu svůj vlastní model

Sepište zdůvodnění, proč jste zařadili dané proměnné

Hodnocení modelů – pravidla
(% správně predikovaných / % použitých proměnných)

Vložte do odevzdávárny (i se jmény a UČO autorů)

45-60 minut včetně přestávky

ZPRACOVÁNÍ A ANALÝZA DAT

Používáte nějaký statistický software?

Jak si připravujete data?

Jaké statistické metody používáte?

Ukázka SPSS Modeler – „overfitting“ na neuronových sítích

INTERPRETACE STATISTIK

Jak interpretovat statistiky

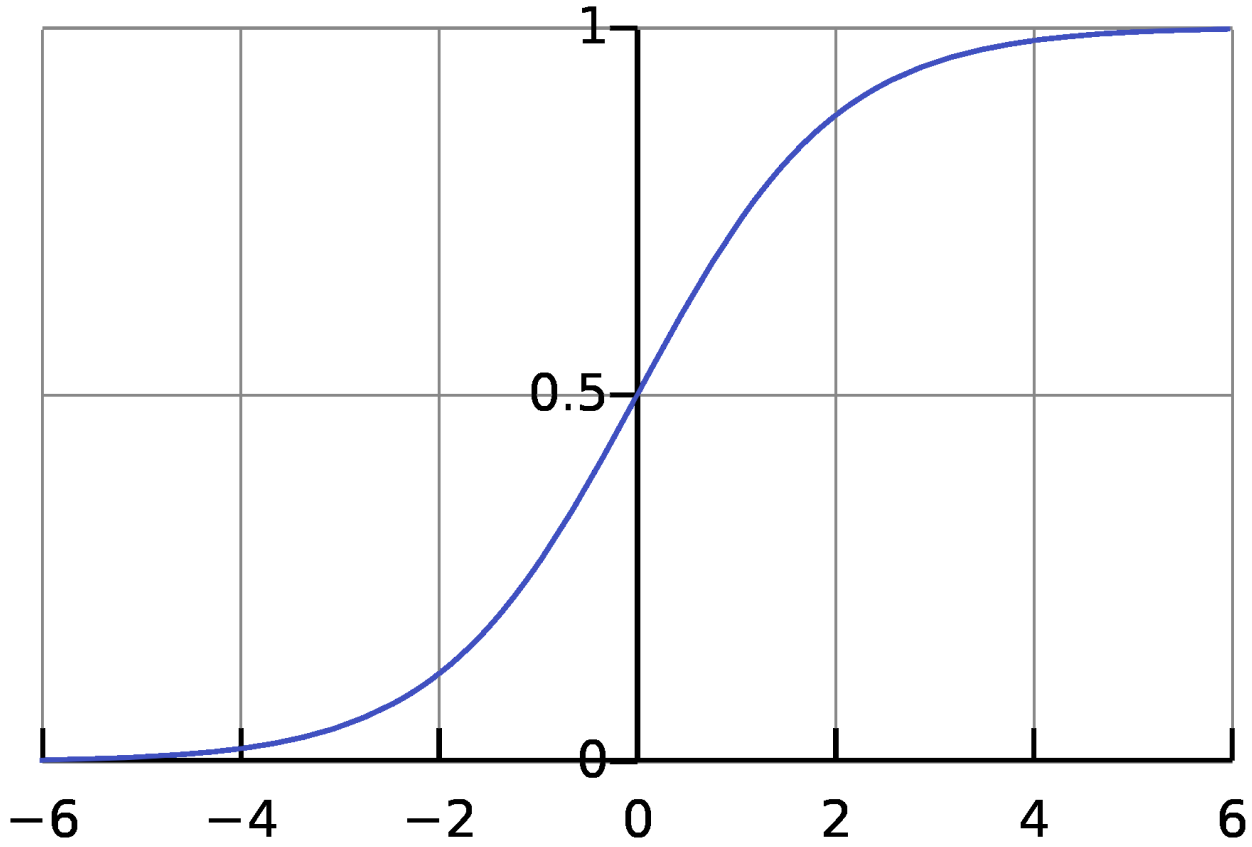
Významnost („Significance“) vs velikost efektu („Effect size“)

Podstata logistické regrese

LOGISTICKÁ RE

[redacted]

1	-1,24	0,29	0,22
[redacted]			
3	-1,74	0,18	0,15
[redacted]			



Parameter Estimates

Attrition[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B)	
								Lower Bound	Upper Bound
Yes	Intercept	-,990	,102	94,780	1	,000			
	JobSatisfaction	-,249	,037	45,796	1	,000	,779	,725	,838

a. The reference category is: No.

VÝPOČET MODELŮ

Zpětná vazba (prosím o vyplnění): bit.ly/PIPMfdb

Vaše výsledky

Co si odnést?