Analýza dat

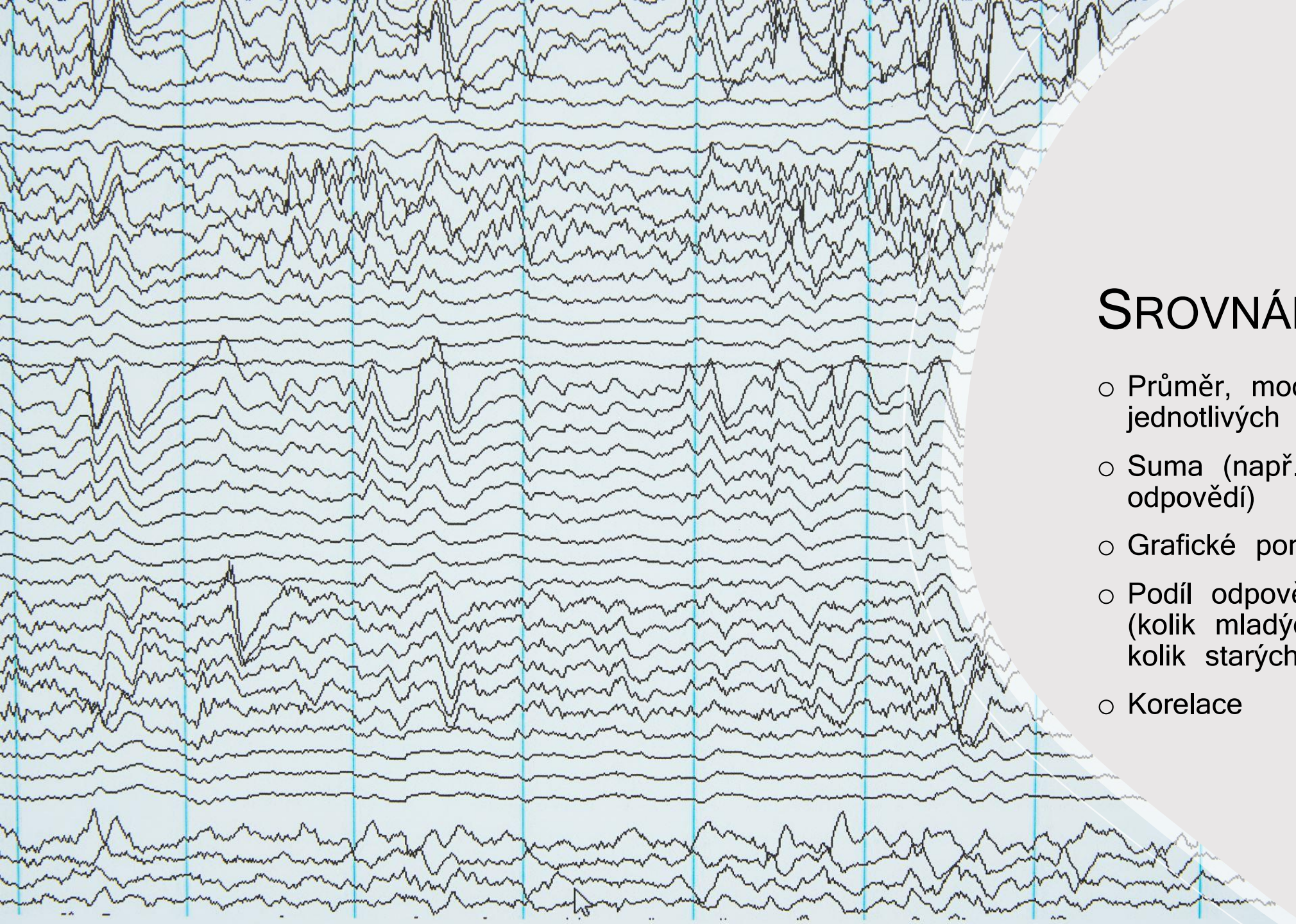
Karel Němeček,
471344@mail.muni.cz

V KOSTCE:

- Shrnuji data → potom srovnávám něco s něčím + měřím vliv něčeho na něco
- Pod vlivem výzkumných otázek
- Vážu analýzu na ověřování hypotéz

DATASET

- Neboli datová matice
- Tabulka, kde jsou záznamy ze všech dotazníků
- **SLOUPCE = PROMĚNNÉ**
- **ŘÁDKY = PŘÍPADY** (zodpovězené dotazníky)



SROVNÁNÍ DAT

- Průměr, modus, medián podle jednotlivých kategorií
- Suma (např. počet správných odpovědí)
- Grafické porovnání
- Podíl odpovědí podle rysu (kolik mladých volí stranu X a kolik starých?)
- Korelace

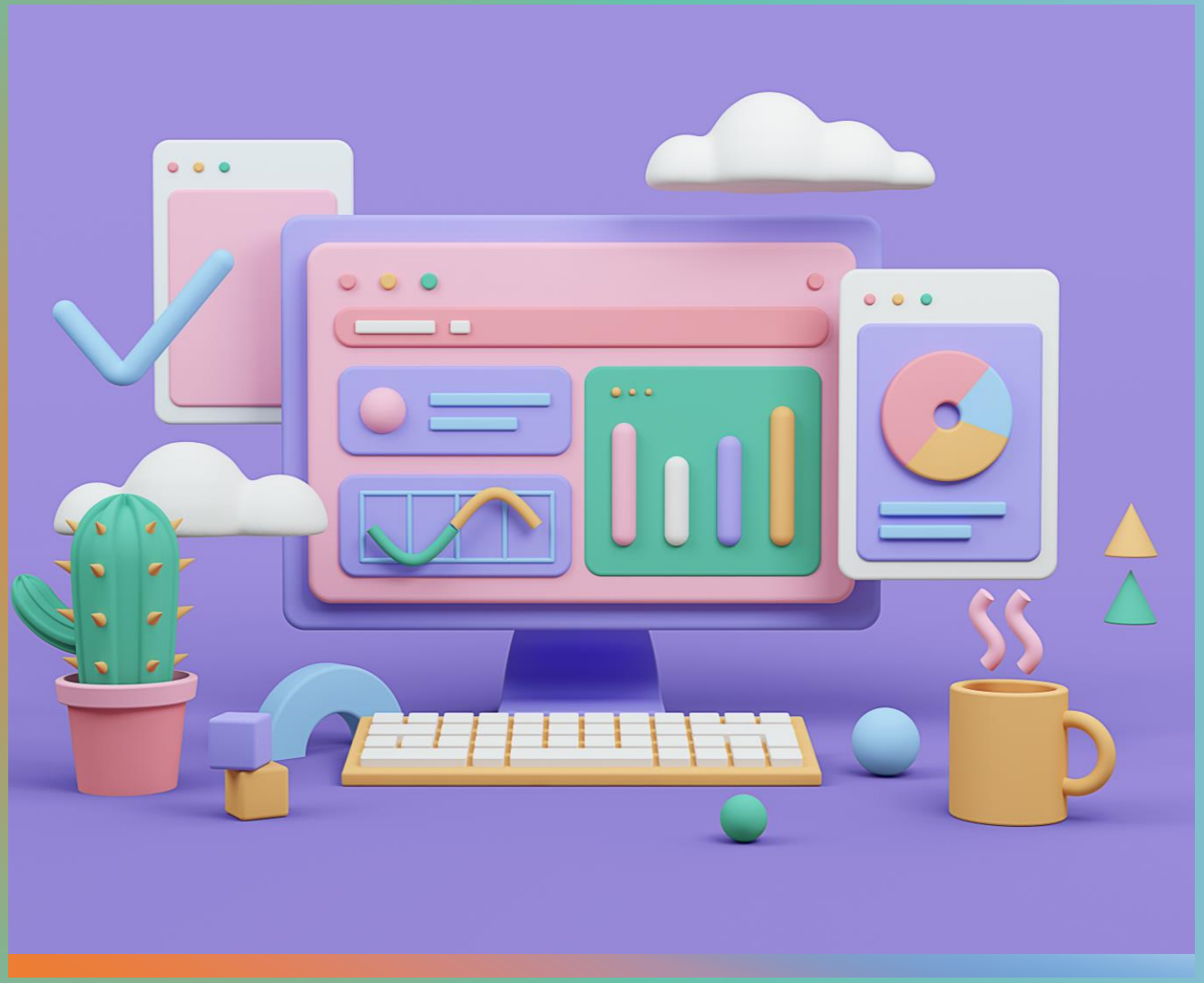
Pár řešení: podle úrovně měření

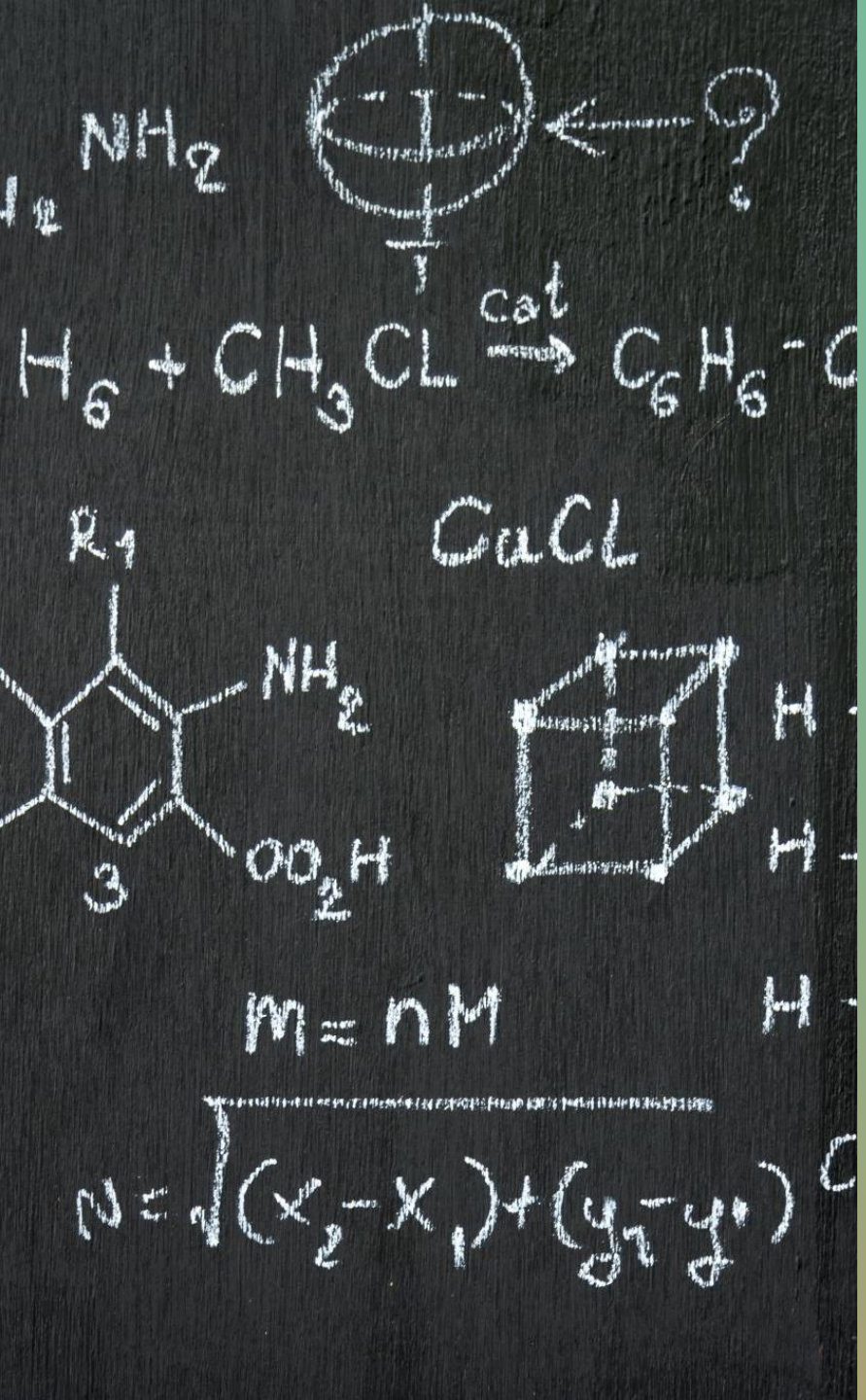
POPIS

- **Málo kategorií** = frekvenční tabulka
- **Mnoho kategorií** = střední hodnoty (průměr atd.)

SROVNÁNÍ

- **Nominální s nominální** → crosstabs
- **Nominální s kardinální** → srovnání průměrů
- **Kardinální s kardinální** → korelace





Použití

- Filtr = pracuji jen s řádky s vybranou odpovědí v sloupci
- Podíl (procenta) = (typ odpovědi/všechny odpovědi)*100
- Průměr = součet sloupce/všechny odpovědi ve sloupci
  - =AVERAGE, =AVERAGEIF
- Median = seřadit od min do max -> potom hodnota na pozici dělící na 2 poloviny nebo hodnota mezi 2 prostředními
  - =MEDIAN
- Modus = nejčastěji se objevující hodnota
  - =MODE

KONTINGENČNÍ TABULKA

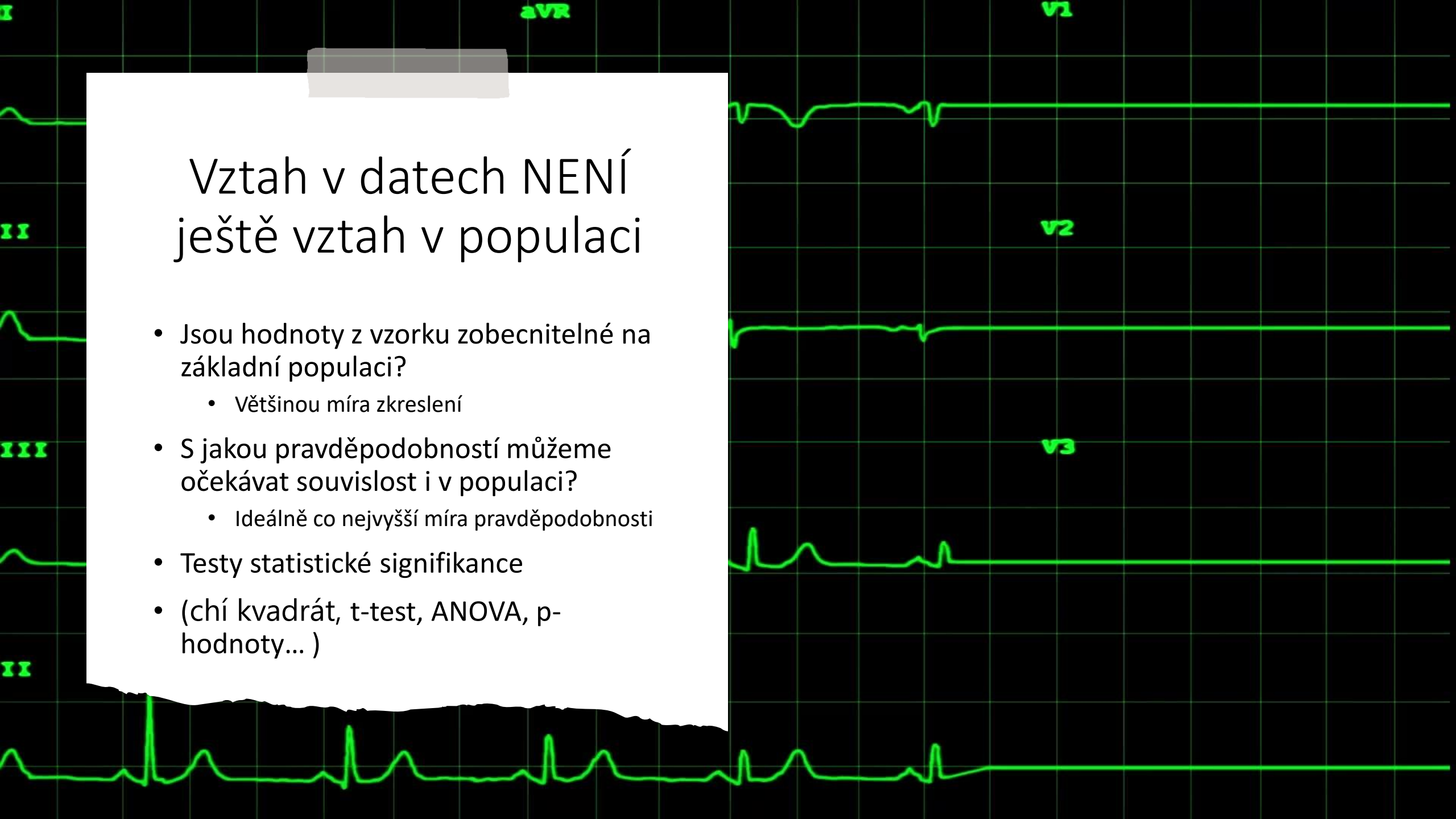
- Anglicky crosstabs
- Tabulka ukazující podíly, četnosti nebo například průměry podle kategorií 2 (a více) proměnných
- PŘÍKLAD: N: ročník na SŠ, Z: zájem o politiku

	Velmi	Trochu	Vůbec	Celkem
Prvák	5 %	24 %	71 %	100 %
Druhák	12 %	35 %	53 %	100 %
Třeťák	18 %	37 %	45 %	100 %
Čtvrťák	24 %	52 %	24 %	100 %

Korelace

- Nutné použít různé typy podle typu proměnné -> jiné výpočty
  - Pro jednoduchost: Pearsonův korelační koeficient
  - =CORREL
- Výsledek: od -1 (nepřímá úměrnost) do 1 (přímá uměrnost)
  - Výsledek 0 = žádná souvislost
  - Používat ideálně na srovnání kardinální s kardinální (číslo o mnoha kategoriích, dají se řadit, jasné měřítko (o kolik větší?))

Hodnota korelace	interpretace souvislosti
0,01 – 0,09	triviální, žádná
0,10 – 0,29	nízká až střední
0,30 – 0,49	střední až podstatná
0,50 – 0,69	podstatná až velmi silná
0,70 – 0,89	velmi silná
0,90 – 0,99	téměř perfektní



Vztah v datech NENÍ ještě vztah v populaci

- Jsou hodnoty z vzorku zobecnitelné na základní populaci?
  - Většinou míra zkreslení
- S jakou pravděpodobností můžeme očekávat souvislost i v populaci?
  - Ideálně co nejvyšší míra pravděpodobnosti
- Testy statistické signifikance
- (chí kvadrát, t-test, ANOVA, p-hodnoty…)

Praktická ukázka?