

kapitola 13

Kde lžou data,
když čísla sedí?

Díry, které tam nebyly

1943, Kolumbijská univerzita. Abraham Wald počítá zranitelnost letadel — jenže díry lze sečíst jen na strojích, které se **vrátily**.

Málo zásahů na motorech neznamená, že se do nich netrefuje.
Znamená to, že zásah do motoru se domů nevrací.

Slavný mem s tečkami je z roku 2016. Wald nenapsal bonmot — napsal metodu.

Data nelhala čísla

- Každá díra na každém vráceném letadle byla spočítána **správně**.
- Data lhala **vzorkem** — tím, kdo se do tabulky vůbec směl dostat.
- Síto pracovalo proti směru otázky: čím pečlivěji jsi počítal, tím jistěji tě vedlo pancéřovat nesprávná místa.
- Čísla seděla. **Právě proto byla nebezpečná.**

Jeden mechanismus, čtyři kostýmy

- Klam přeživších — počítáš jen letadla, která se vrátila.
- Paradox porodní váhy — cigareta „chrání“ lehké děti (30 let záhady).
- Vždy totéž: podmínění na následku.
Podmínka výběru je tichý collider.

Tužka před regresí

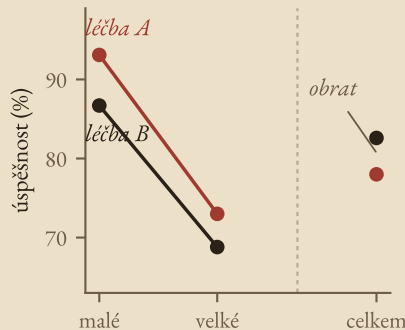
Obrana nestojí větší data — stojí **dvacet vteřin tužkou**.

1. Kdo ve vzorku chybí? 2. Co bylo podmínkou vstupu? 3. Kdo měřil a proč?

Nakresli kauzální graf — pak stratifikuj podle **příčin**, ne podle následků.

Regrese bez grafu je výpočet bez otázky.

Simpsonův obrat: aritmetika vah



Léčba A vede v obou vrstvách — a v souhrnu prohrává. Obrat nevěží v datech, ale ve vahách.

„Jak bych poznal, že se pletu?„

Až ti příště analýza potvrdí přesně to, co sis přál, vezmi tužku dřív než šampaňské.

A ptej se s Waldem: kde jsou letadla, která se nevrátila?