



StatSoft

# Pohádkový soudní proces

Pojďme si nyní vysvětlit princip testování hypotéz na příkladu se středověkým soudním procesem. A v duchu celého vánočního vydání na Vás bude čekat další pohádková postavička. Nejprve však trocha teorie, abychom se v příběhu neztratili.

Dosud jsme si v našem seriálu ukazovali aplikace výpočtů, které svým charakterem spadají do deskriptivní statistiky. Vypočtené charakteristiky polohy a variability doplněné patřičnými grafy nám dají velmi slušný přehled o našem konkrétním datovém souboru, nicméně občas bychom rádi například rozhodli o správnosti nějakého tvrzení (střední hodnota pro danou veličinu je 0, veličiny jsou nezávislé, data pocházejí z normálního rozdělení,...). Chtěli bychom mít nějaký prostředek, který by rozhodl, jak to vlastně je. Takovéto otázky řeší oblast testování hypotéz, kterou si dnes opět odlehčenější formou představíme.

## Postup testování hypotéz

Jaký je postup testování hypotéz? Začít bychom měli tím, že máme nějakou tezi, domněnku, tvrzení. Náš výrok potom chceme prověřit pomocí naměřených dat. Otázkami, které nás mohou zajímat, mohou být například:

- › Jsou muži v průměru vyšší než ženy?
- › Mají studenti žijící v Praze méně sourozenců než ostatní studenti?
- › Je chirurgická léčba lepší než léčba medikamentózní?
- › Je nový lék účinnější než starý?
- › Jak často mám provádět seřízení stroje?
- › Je IQ žáků deváté třídy vyšší než 100?

K tomu, abychom mohli prokázat platnost našeho tvrzení, potřebujeme dvě věci: Někakého arbitra, který rozhodne o platnosti či neplatnosti našeho výroku a důkazní materiál, které by ho přesvědčil. Arbitrem nebude v našem případě člověk, ale statistický test. Důkazním materiálem nebudou otisky prstů, nahrávky, kousky vlasů s DNA na místě činu, ale data. Náš výběrový soubor dat by pro účely daného testu měl splňovat určité předpoklady, o kterých si ale povíme někdy příště. Dnes nám jde pouze o princip testování hypotéz.

# Základní pojmy

Než se podíváme na slíbený pohádkový příběh z dávné minulosti, definujeme si několik základních pojmů:

**Testování hypotéz** – ověřování platnosti nějakého výroku (tvrzení).

**Statistický test** – pravidlo, které rozhoduje o platnosti nebo neplatnosti naší hypotézy. Kritériem bývá většinou velikost nějaké funkce, která je spočítána přímo z datového souboru, při různých datech nabývá různých hodnot (je to tedy náhodná veličina).

**Testová statistika** – funkce, pomocí které rozhodujeme.

Je vždy nutné zformulovat dvě hypotézy:

**$H_0$ : Nulová hypotéza**, o které se primárně předpokládá, že platí. Testová statistika se tedy řídí rozdělením, jako by platila nulová hypotéza.

**$H_1$ : Alternativní hypotéza** (někdy také označováno jako  $H_A$ ), která platí, pokud je zamítnuta hypotéza nulová.

Pro konstrukci testu je potřeba definovat obě hypotézy, nestačí jen  $H_0$ , musíme přeci vědět, co platí, když  $H_0$  neplatí.

Poznámka: V praxi se jako alternativní hypotéza definuje obvykle náš výzkumný záměr, kvůli kterému jsme začali sbírat data a zahájili výzkumný projekt. Sbíráme data, provádíme výzkum a poté už jen můžeme doufat, že test zamítne nulovou hypotézu a my tedy prokážeme náš záměr (typicky: *chceme ukázat, že lék má pozitivní vliv na pacientovo zdraví*,  $H_0$  tedy je, že *nemá pozitivní vliv*,  $H_1$  že *má pozitivní vliv*).

Výsledkem testu může být zamítnutí  $H_0$  nebo nezamítnutí  $H_0$ .

Z filozofického hlediska platí u testu, stejně jako u soudního řízení, presumpce nevin, tedy nulová hypotéza platí (testová statistika se řídí rozdělením podle  $H_0$ ), pokud jí nezamítneme. Hypotézy u soudního procesu mají tuto podobu:

$H_0$ : *Obžalovaný je nevinný*

$H_1$ : *Obžalovaný je vinen*

Zde se však bijí dva či dokonce tři úhly pohledu. Pohled soudce, který ctí právě presumpci nevin. Pohled žalobce, který je přesvědčen o opaku, tedy o platnosti alternativní hypotézy. I když by soudce odsouzeného nevydal, stejně bude trvat na tom, že je vinen, ale tentokrát pouze neměl dost důkazů k jeho prokázání. A to ještě nemluvíme o obhájci ☺. Kdybychom to zjednodušili, soudce je náš test, žalobce je ten, kdo sbírá data vhodná pro zamítnutí hypotézy, obhájce zase sbírá data pro podpoření hypotézy nulové.

Na první pohled se může zdát, že máme pouze dva scénáře, které mohou nastat, zjednodušeně řečeno: vinen/nevinen. Stejně jako soudce, tak i statistický test se na základě předložených důkazů, resp. dat, může splést. Tedy může prokázat tvrzení, které ve skutečnosti neplatí. Kromě správného verdiktu mohou nastat další dvě situace:

**Nulová hypotéza ve skutečnosti platí, ale test prokázal alternativu**

**Alternativní hypotéza ve skutečnosti platí, ale test nezamítl nulovou**

Ve statistice se tyto varianty pochybení nazývají:

**Chyba I. druhu** (její velikost se běžně označuje jako  $\alpha$ )

**Chyba II. druhu** ( $\beta$ )

Různé možnosti vztahu mezi skutečností a rozhodnutím soudu (výsledkem testu) ukazuje následující tabulka:

		rozhodnutí	
		obžalovaný je nevinen ( $H_0$ platí)	obžalovaný je vinen (platí $H_1$ )
skutečnost	obžalovaný je nevinen	správné rozhodnutí	chyba I. druhu
	obžalovaný je vinen	chyba II. druhu	správné rozhodnutí

Pravděpodobnost chyby prvního druhu ( $\alpha$ ) nazýváme **hladina významnosti testu**.

Pravděpodobnost správného zamítnutí neboli **síla testu** je  $1-\beta$ .

Tak to by byly pojmy, nyní si princip testování hypotéz ilustrujeme na čarodějnickém soudním procesu.

## Čarodějnický soudní proces

Na úvod se sluší říci, že podobnost s jinými příběhy či reálnými událostmi je čistě náhodná.

Tento příběh se podle dostupných zdrojů odehrál již před mnoha lety v místě, myslím, zvaném Středozemě. Jeden z místních obyvatel, který se posléze stane obviněným v tomto příběhu, si všiml podobných znaků mezi synem a otcem, matkou a dcerou. A tak začal tyto znaky předpovídat přímo i u nově narozených dětí. Jeho (přesněji řečeno její) předpovědi se po narození s postupem času, tak jak miminka stárá, vyplňovaly. Po té, co se s odstupem času potvrdilo několik relativně úspěšných předpovědí, byla tato osoba, mladá dáma, obviněna z čarodějnictví. Zajímavé, že její kamarád, se kterým na tomto výzkumu také pracovala, obviněn nebyl. Soud středověkého charakteru, který je v danou dobu považován za jediný spravedlivý, zasedl.

Jde o jednoduchý soudní proces, ve kterém existuje pouze jediný možný trest (upálení). Soud rozhodne, jestli se tomuto trestu obžalovaná podrobí anebo nepodrobí. Na rozdíl od dnešních soudů zde neexistuje žádné odvolání. Obžalovaná je obviněna z čarodějnictví, což je velmi závažné obvinění. Mimochodem, kdyby byla například obviněna z toho, že je pohádková bytost, tak by to ještě šlo, neboť vyhoštění z vesnice do pohádkového lesa se dá přežít, protože v lese důsledkem soudních procesů mezitím vyrostlo vlastní, ve stromech skryté město, ve kterém žijí všichni za pohádkové bytosti prohlášení. Byl zde skoro každý člověk, co v něčem (kromě hrubé síly) výrazně vynikal.

No ale čarodějnictví, to je horší. „Můžou nastat dvě situace, jak to dopadne. Buď prokážeme, na základě důkazů z posledních let, že je to čarodějnice, anebo nám tyto důkazy stačit nebudou, ale čarodějnice to je tak i tak,“ sdělí žalobce den před procesem svému kolegovi. Pokud je test statistiky významný, pak žalobce prokáže svou tezi (alternativní hypotézu). Zamítnutí nulové hypotézy je ekvivalentní přijetí alternativy. Pokud však test nezamítne, žalobce zkoumanou



hypotézu neprokáže, ale to neznamená, že neměl pravdu. Nemyslete si, pan žalobce si bude pořád myslet, že má pravdu, třeba jen neměl dostatek důkazů.

Jak už jsme si naznačili výše, mohou nastat dva případy špatného rozhodnutí:

- I. **Obžalovaná je nevinná, ale soud jí odsoudí a pošle na hranici.**
- II. **Obžalovaná je vinná, je to skutečně čarodějnice, ale důkazní materiál předložený žalobou na to nestačí a odsouzena a upálena na hranici nebude.**

Které z těchto pochybení je horší? Neupálit nebezpečnou pohádkovou bytost oplývající magickými schopnostmi nebo upálit nevinnou ženu, které nemá s kouzly nic společného? Odpověď si řekneme za chvíli, nyní se pojďme podívat, jak tento příběh dopadl.

## Soudný den

Pan soudce, který tento proces soudí, je relativně liberální a přičítá se mu, že by odsoudil nevinného člověka. Je známý tím, že potřebuje dostatečné množství důkazů na to, aby prokázal tvrzení žalobce, tedy rozhodl o vině. V této historické době nevídaná věc, neboť většina soudců je velmi konzervativních až posedlých a k odsouzení jim stačí i nepřímé a často hloupé důkazy. Je pravda, že skutečný viník jim nikdy neutekl, ale na hranicích a v žalářích končí velké množství populace úplně zbytečně. Potom se není čemu divit, že v pohádkovém lese vyrostlo již zmíněné město odsouzených, složené z převážně nadaných lidí.



Souzená má tedy štěstí, že bude soudit tento soudce (mezi kolegy pro svůj styl nazývaný „Měkouš“). V naší terminologii můžeme tohoto soudce přirovnat k testu s velmi nízkou chybou prvního druhu (chce mít jistotu, že neodsoudí nevinného) nebo například k neparametrickému testu, ale o tom v některém z budoucích čísel. U tohoto soudce se ale může stát a často se to i stává, že obvinění zamítne, ve skutečnosti je však viník „zlosyn“. Žalobce prostě nesehnal dostatek důkazního materiálu k přesvědčení soudce. Soudcovi drsnější kolegové bychom poté přirovnali naopak k testům s velkou hodnotou  $\alpha$ .

## Obžalovaná povstaňte

Podle dochovaných záznamů v šuplíku v kanceláři StatSoftu se dochovala část přepisu tohoto dějství:

*Soudce:* „Stojíte před vážným obviněním, jste obviněna z čarodějnictví. Prosím pana žalobce o předložení důkazů.“

*Žalobce:* „Důkazy jsou zde,“ ukáže na syna s blondými vlasy a modrými očima, které jsou stejné, jako má jeho otec. „Obžalovaná rodině sama řekla, že syn bude mít tyto rysy a přitom se ještě nenarodil. Není to náhoda, stejné je to u těchto dalších 4 dětí“.

*Soudce:* „Tvrdíte, že předpověděla barvu očí a vlasů u všech těchto dětí, ještě před jejich narozením?“

*Žalobce:* „Ano, a u těchto dvou dokonce jejich výšku,“ řekl žalobce a ukázal na enormně vysokého chlapce, jehož otec mezitím zemřel, ale patřil k nejvyšším dlouhánům ve vesnici.

*Soudce:* „Hmm, tyto důkazy jsou celkem závažné, nechť promluví obžalovaná.“

*Obžalovaná:* „Pane soudce, všimla jsem si, že těmito znaky existuje souvislost, dědičnost, tedy, že existuje závislost mezi výškou otce a výškou jeho syna.“

Matka obžalované po tomto výroku omdlela.

*Soudce: „V kolika případech se obžalovaná spletla při svých předpovědích?“*

*Obhájce: „Celkem u zhruba dalších pěti dětí, kde předpověděla úplně jiné znaky, než má jejich otec. Je to tedy celé pouhá spekulace. Trváme na tom, že obžalovaná je člověk a vždy jím byla.“*

*Soudce: „Aha, takže ono to zas s tou předpovědí není tak horké, s podobným úspěchem předpovídáme počasí, vezmu si čas na rozmyšlenou...“*

Po chvíli soudce vyřkl rozsudek.

*Soudce: „Obžalovaná je zproštěna obvinění, tvrzení, že je čarodějnice, se neprokázalo. Stále platí, že je člověk jako každý z nás v této místnosti a tak s ní budeme i nakládat.“*

Obžalovaná byla tedy pro nedostatek důkazů propuštěna, ale bylo to správně? Zachránilo jí, že její předpovědi měly přibližně 50% úspěšnost, ale nebylo to pouze nevhodným vzorkem dat? Tento vzorek byl velmi malý. Zsvěcení navíc věděli, že obžalovaná měla pravdu i v mnoha dalších případech, ale protože si opravdu mysleli, že je čarodějnice, tak ze strachu raději mlčeli a o skutečném původu některých dětí se raději nezmiňovali. V městečku tedy panovala podivná atmosféra a zhruba půlka stále věřila, že se o bytost s nadpřirozenými schopnostmi opravdu jedná. Proto někteří začínali pomalu shánět další důkazy pro nový soudní proces, tedy nový vzorek dat.

## Shrnutí

Ať už se soudce mýlil nebo ne, v důkazním materiálu bylo příliš mnoho pozorování, která hovořila v neprospěch alternativy, kterou chtěla obžaloba prokázat.

V úvodu jsme vymezili dva případy, jak se může soudce zmýlit. Která z těchto pochybení je závažnější? Všeobecně se domníváme, že mnohem závažnější je odsoudit nevinného člověka než neodsoudit skutečného zločince (navíc, upálíme-li někoho, jistě s tím později již nebudeme moci nic dělat) a takto se chovají i testy. Proto se také běžně stanovuje  $\alpha$  pevně na nějakou hodnotu – chceme mít jistotu pro velikost chyby, že odsoudíme nevinného.

A co soudcovo pojmenování „Měkouš“? Byl tento soudce špatný? Kvalita soudce by měla být zkoumána nejen podle toho, jakou chce mít jistotu neodsouzení nevinného, ale také podle toho, jak správně odsoudí skutečně vinného (což určuje v terminologii testování hypotéz vlastně síla testu). Žádné z těchto pochybení není možné úplně vyrušit, protože sníží-li možnost výskytu jednoho typu pochybení soudce, enormně vzroste výskyt druhého pochybení. Chyby prvního a druhého druhu jdou proti sobě.

Jak tedy poznat kvalitního soudce? V teorii testování hypotéz se to řeší následovně: stanoví se pevná hladina  $\alpha$  a z testů, které ji dosahují, se vybere ten, který má nejmenší chybu druhého druhu, tedy největší sílu testu. Rozhodnutí o pojmenování „Měkouš“ bychom tedy museli rozhodnout na základě porovnání s jinými soudci, kteří se chovají stejně liberálně jako on.

*K aplikaci těchto myšlenek se vrátíme v nějakém z příštích čísel.*

## Závěrem

Ať už se příběh z našeho zaprášeného šuplíku stal nebo nestal, osobně jsme s rozsudkem soudce spokojeni. Přeci jenom by nám bylo líto odsouzení nadějně mladé statističky z čarodějnictví. Popravdě jsme si i trochu oddechli, můžeme totiž dále bezpečně chodit do práce bez vidiny upálení. Prozatím.