

## Kapitola 2 - Testování hypotéz. Testy dobré shody

Dva základní statistické postupy jsou odhad parametrů a testování hypotéz. V minulé kapitole jsme si ukázali, jak odhadujeme charakteristiky základního souboru, v této kapitole probereme základy testování hypotéz (*hypothesis testing*). Mezi základní poučky metodologie vědy patří, že shoda dat s hypotézou ještě neznamená, že hypotéza je pravdivá; naproti tomu data odporující hypotéze ukazují na to, že hypotéza pravdivá není. Hypotézu nelze na základě dat dokázat; hypotézu však lze na základě dat vyvrátit. z toho vychází i statistické testování hypotéz. Ukážeme si jej na příkladě vyhodnocení nominálních dat; ne proto, že by se pro jiné typy dat neužívalo, ale proto, že je na nominálních datech nejsnáze pochopitelné. Postup je následující:

Formuluji nulovou hypotézu. Nulová hypotéza je formulována tak, aby ji mohla data vyvrátit v případě, že není pravdivá. Většinou to tedy bývá opak toho, co chci dokázat. Nulová hypotéza (*null hypothesis, H<sub>0</sub>*) je většinou formulována jako: něco se neliší; není diference; není závislost; platí zákon atd. Poté se snažím dokázat, že určitá data nejsou slučitelná (jsou v rozporu) s touto nulovou hypotézou. Pokud to dokáži, zamítám nulovou hypotézu a přijímám alternativní hypotézu  $H_A$ , někdy též  $H_1$ , která je negací nulové hypotézy.

*Příklad:* Studuji dědičnost rostliny a ptám se, zda zde platí pro barvu květů jednoduchá mendelovská dědičnost. Předpokládám, že v  $F_1$  generaci bude poměr počtu červenokvětých k bělokvětým 3:1. Mám 80 potomků. Potom předpokládám, že v potomstvu bude 60 červenokvětých a 20 bělokvětých jedinců. Já však mám 10 bělokvětých a 70 červenokvětých. Jsou moje výsledky v rozporu s poměrem 3:1, tzn. s tím, že každé individuum má pravděpodobnost 0.75 být červenokvěté a 0.25 být bělokvěté?

I v případě, že pravděpodobnosti jsou 0.75 a 0.25, mohu s určitou pravděpodobností dostat výsledný poměr 70:10. Dokonce mohu dostat všech 80 červenokvětých (s pravděpodobností  $0.75^{80}$ , což je řádově  $10^{-10}$ ). V takovém případě ovšem nebudu ochoten věřit, že šlo pouze o náhodu, a dojdu k názoru, že nulová hypotéza neplatí (zamítну ji). **Nulovou hypotézu zamítám, pokud dostanu uspořádání dat, které je za předpokladu platnosti nulové hypotézy velmi nepravděpodobné.**

Co to ale je, velmi nepravděpodobné? Statistika nám k tomu poskytuje následující návod: Zvol si, jak nepravděpodobný výsledek za předpokladu platnosti nulové musíš dostat, abys ses rozhodl pro závěr, že nulová hypotéza neplatí. Většinou se rozhodujeme pro 5% nebo 1%. Této hodnotě říkáme hladina významnosti testu a značíme ji  $\alpha$ ; bývá zvykem ji vyjadřovat desetinným číslem, např.  $\alpha=0.05$ . Potom spočti testové kritérium (někdy tuto hodnotu nazýváme testová statistika; zde je další význam termínu statistika). Pro toto kritérium je známo, jaké má rozdělení v případě platnosti nulové hypotézy. Je tedy známo, kterou hodnotu překročí s pravděpodobností 5%, kterou hodnotu překročí s pravděpodobností 1% atd. Těmto hodnotám říkáme kritické hodnoty. Jestliže hodnota testového kritéria překročí kritickou hodnotu pro zvolenou hladinu významnosti, zamítne nulovou hypotézu na dané hladině významnosti. Říkáme potom, že výsledek (nesouhlas s nulovou hypotézou) je signifikantní na dané hladině významnosti. Pro kategoriální data používáme testy dobré shody (*goodness of fit*) a používáme kritéria (statistiky)  $\chi^2$ , čti

„chí-kvadrát“, angl. *chi-square*, čti „kaj“ (příšerně působí běžně používaná česko-anglická směs, vyslovovaná „chí-skvér“):

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - \hat{f}_i)^2}{\hat{f}_i}$$

### Vz. Kapitola 2 -A

$k$  je celkový počet kategorií, které sleduji (v našem příkladě 2),  $\hat{f}_i$  je očekávaná četnost v  $i$ -té kategorii (často se také značí  $E$ , z anglického *Expected*,  $f_i$  je četnost skutečná (pozorovaná, někdy též  $O$ , *Observed*). V našem případě tedy formuluji nulovou hypotézu: v  $F_1$  generaci je poměr pravděpodobností výskytu červenokvětých k bělokvětým 3:1; v 80-tičlenném potomstvu tedy předpokládám 60 a 20 individuí.

$$\chi^2 = \frac{(70 - 60)^2}{60} + \frac{(10 - 20)^2}{20} = 6.66$$

Tuto hodnotu porovnávám s tabulkovou hodnotou kritických hodnot  $\chi^2$  pro danou hladinu významnosti  $\alpha$  (většinou volíme 0.05 nebo 0.01) a daný počet stupňů volnosti. Stupně volnosti, (*degrees of freedom*; srovnej se zvoláním bojovníků za práva černochoů, vloženým do spirituálu: *Oh, Freedom*), značíme je většinou d.f., DF nebo  $\nu$  („ný“). Pro testy tohoto typu je počet stupňů volnosti roven počtu kategorií zmenšenému o jednu ( $k - 1$ ). Je to počet četností ve skupinách, které potřebujeme znát, abychom znali celý výsledek. Počet případů v poslední kategorii můžeme dopočítat ze znalosti předcházejících  $k-1$  kategorií a celkového počtu pozorování (ten je v testech považován za fixní). Ví-li, že z osmdesáti jedinců bylo sedmdesát červenokvětých, znám výsledek celého pokusu. Kritické hodnoty najdeme ve statistických tabulkách (viz tabulka kapitola 2 -c); dnes jsou tyto hodnoty ve většině statistických programů, takže se běžný uživatel již s tabulkami většinou nesetká. Získaná hodnota 6.66 je větší než kritická hodnota  $\chi^2_{0.05,1}$ , (tj. při 5% hladině významnosti a jednom stupni volnosti), jejíž hodnota je 3.84. Zamítáme tedy nulovou hypotézu na pětiprocentní hladině významnosti. Závěr by tedy zněl: Pozorovaná data se významně (signifikantně) na 5%-ní hladině významnosti liší od četností, předpokládaných jednoduchou mendelovskou dědičností. V našem případě by odlišnost byla průkazná i na 1%-ní hladině významnosti.

Pozor: v testu užíváme přímo napozorované četnosti. Nelze převést nejprve údaje na procenta a potom počítat s procenty!!!

*Příklad:* V jeskyni je velké množství netopýrů (pro nás jich je nekonečně mnoho), samci a samice. Chci zjistit, zda je poměr samců a samic 1:1. Nejsem ale schopen prohlédnout všechny netopýry v jeskyni. Chytím jich tedy 100 a podle nich se snažím rozhodnout. Oněch 100 individuí musí být náhodným výběrem! Nulová hypotéza zní: V jeskyni je stejně samců jako samic (což je totéž jako: pravděpodobnost, že náhodně vybrané individuum je samec, je stejná, jako že náhodně vybrané individuum je samice).

Existují dvě možnosti, jak je tomu ve skutečnosti:

1. V jeskyni je stejně samců jako samic, obě pravděpodobnosti jsou tedy 0.5. To znamená, že nulová hypotéza platí (je pravdivá). Výsledek pokusu může být dvojí:

1a) např. 55 samců; 45 samic. Potom  $\chi^2=(55-50)^2/50+(45-50)^2/50 = 1.0 < 3.84$ . Nemohu zamítnout nulovou hypotézu. Správné rozhodnutí.

1b) např. 60 samců; 40 samic. Potom  $\chi^2=(60-50)^2/50+(40-50)^2/50 = 4.0 > 3.84$ . Zamítám nulovou hypotézu na 5%-ní hladině významnosti. Udělal jsem **chybu prvního druhu - Type I error**. Pravděpodobnost této chyby známe: je to  $\alpha$ . Hladina významnosti  $\alpha$  je tedy podmíněná pravděpodobnost zamítnutí nulové hypotézy za předpokladu, že nulová hypotéza platí.

2. Samci tvoří 60% individuí, náhodně vybrané individuum bude samec s pravděpodobností 0.6; samice s pravděpodobností 0.4. Nulová hypotéza tedy neplatí - je nepravdivá. Výsledek pokusu může být opět dvojí:

2a) např. 55 samců; 45 samic. Potom  $\chi^2=(55-50)^2/50+(45-50)^2/50 = 1.0 < 3.84$ . Nemohu zamítnout nulovou hypotézu. Dopustil jsem se **chyby druhého druhu**. Její pravděpodobnost označujeme jako  $\beta$  a většinou ji neznáme.  $1 - \beta$  je **síla testu (power of the test)**. Obecně platí, že síla testu roste s odchylkou od nulové hypotézy a s počtem pozorování. Dále platí, že čím menší je  $\alpha$ , tím větší je  $\beta$ . Protože  $\beta$  neznáme, je správná formulace výsledku: **Na základě dat nemůžeme zamítnout nulovou hypotézu**. Formulace: *Dokázali jsme nulovou hypotézu je nesprávná!*

2b) např. 60 samců; 40 samic. Potom  $\chi^2=(60-50)^2/50+(40-50)^2/50 = 4.0 > 3.84$ . Zamítám nulovou hypotézu na 5%-ní hladině významnosti. Správné rozhodnutí.

Máme tedy dvě možnosti, jaká je realita (nulová hypotéza buď platí nebo neplatí) a naše rozhodnutí může být také dvojí (nulovou hypotézu zamítám, nebo nezamítám). Celý proces je zvykem ilustrovat tabulkou:

Naše rozhodnutí	Skutečnost	
	Je-li $H_0$ pravdivá	Je-li $H_0$ nepravdivá
$H_0$ jsme zamítli	Chyba 1. druhu	Správné rozhodnutí
$H_0$ jsme nezamítli	Správné rozhodnutí	Chyba 2. druhu

Tab. Kapitola 2 -A Chyba 1. a 2. druhu při statistickém rozhodování

Chyba prvního i druhého druhu jsou vlastní statistickému rozhodování a vyplývají ze stochastického (náhodného) charakteru studovaných procesů; nelze je tedy žádným způsobem z našeho rozhodování zcela eliminovat. Čím menší pravděpodobnost chyby prvního druhu jsme ochotni připustit, tím větší máme pravděpodobnost chyby druhého druhu. Představme si v příkladu netopýrů, že jsme ochotni přijmout pravděpodobnost chyby prvního druhu pouze 0.01. Kritická hodnota testu je 6.63. Co je toho důsledkem? V případě 1b jsem se díky přísnosti kritéria nedopustil chyby prvního druhu; naproti tomu jsem se v případě 2b dopustil chyby druhého druhu. Za lepší ochranu před chybou prvního druhu platím větší pravděpodobností chyby druhého druhu.

Na tomto příkladě však lze demonstrovat nebezpečí jiných chyb, které sice nejsou vlastní statistice, ale při aplikaci statistických metod na biologické problémy se jim většinou také nevyhneme. Oněch 100 individuí pokládáme za náhodný výběr. Nicméně, abychom opravdu mohli provést náhodný výběr, museli bychom všechna individua očíslovat a potom podle tabulky náhodných čísel vybrat 100 individuí - to

logicky není možné. Za náhodný výběr obvykle považujeme ta individua, která se nám podaří získat. Předpokládejme, že sbíráme netopýry v zimě, když visí ze stropu jeskyně. Pokud si např. samci vybírají pro přezimování pro člověka obtížněji dostupná místa než samice, nebo se samci rychleji probudí a dřív nás pokoušou, takže nám jich víc uletí, je pravděpodobné, že v našem výběru bude (statisticky významně) více samic - nepodařilo se nám provést náhodný výběr. V praxi tedy může být zamítnutí nulové hypotézy důsledkem tří skutečností:

1. Nulová hypotéza neplatí.
2. Nulová hypotéza platí, ale dopustili jsme se chyby 1. druhu.
3. Nulová hypotéza platí, ale my jsme nesplnili všechny předpoklady pro užití testu.

Test a základní vzorec

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - \hat{f}_i)^2}{\hat{f}_i}$$

je možno použít pro libovolný počet kategorií. Následuje příklad (Zar 1984) sledování dvou znaků na semenech (semena zelená, žlutá; svraskalá a hladká). Žlutá a hladká se považují za dominantní. Očekávaný poměr je potom 9:3:3:1. V tomto případě je počet kategorií  $k = 4$ , DF ( $v$ ) = 3. Bylo sledováno 250 semen. Pozorované četnosti fenotypů byly 152, 39, 53, 6.

Nulová hypotéza ( $H_0$ ): Sledovaný výběr semen pochází ze základního souboru charakterizovaného poměrem fenotypů žlutých hladkých, žlutých svraskalých, zelených hladkých a zelených svraskalých 9:3:3:1. (Můžeme též formulovat: pravděpodobnosti výskytu daných fenotypů jsou v poměru 9:3:3:1.) Alternativa ( $H_A$ ): Semena pocházejí ze základního souboru, který nemá poměr shora uvedených fenotypů 9:3:3:1.

Očekávané četnosti spočteme trojčlenkou. Např. očekávaný počet žlutých hladkých semen je  $250 \cdot (9/16) = 140.625$

	žlutá hladká	žlutá svraskalá	zelená hladká	zelená svraskalá	$n$
pozorované	152	39	53	6	250
očekávané	140.625	46.875	46.875	15.625	

Tab. Kapitola 2 -B Příklad užití  $\chi^2$  testu

$$\chi^2 = \frac{11.375^2}{140.625} + \frac{7.875^2}{46.875} + \frac{6.125^2}{46.875} + \frac{9.625^2}{15.625} = 0.9201 + 1.3230 + 0.8003 + 5.9290 = 8.$$

972

Protože 8.972 je větší než kritická hodnota pro  $\alpha=0.05$  (ta je 7.815, viz tabulka 2-3), zamítáme nulovou hypotézu na 5%-ní hladině významnosti. Můžeme tedy zamítnout hypotézu, že data odpovídají modelu jednoduché mendelovské dědičnosti s nezávislými znaky. Z hodnot jednotlivých sčítanců vidíme, že nejvýraznější příspěvek

k vysoké hodnotě testovacího kritéria dává poslední kategorie (semena zelená, svraskalá).

Uvedme další příklady užití tohoto testu:

(1.) Včely jsou postupně vpouštěny do pokusného prostoru se žlutými, červenými a modrými terči. Sledujeme barvu terče, na který každá včela poprvé usedne. Nulová hypotéza je, že pravděpodobnost usednutí nezávisí na barvě terče (tímto způsobem zjišťujeme, zda se včely vizuálně orientují a zda při této orientaci hrají nějakou úlohu barvy). Data: bylo vpuštěno 100 včel; četnosti barev, na které poprvé usedly: žlutá 47, červená 38, modrá 15. Lze z těchto dat usoudit, že včely některou barvu preferovaly? Nulová hypotéza bude znít: Pravděpodobnost usednutí včely na terč nezávisí na barvě terče, a očekávané četnosti budou 33.3 : 33.3 : 33.3. Na tomto pokusu můžeme demonstrovat další podmínky použití tohoto testu: 1. **četnosti pocházejí z nezávislých pokusů.** Proto vpouštíme včely do pokusného prostoru po jedné, a zaznamenáváme chování každé včely. Kdybychom vpustili všechny včely do prostoru najednou a spočetli počet, který se usadil na každém terči, může být (průkazná) odchylka od nulové hypotézy dána tím, že včely poletí společně jako roj, a společně usednou na terč, který náhodně vybere jedna z nich, jakási „vedoucí roje“. Při provádění pokusu si musíme být jisti, že usednuvší včela nenechá na terči nějakou značku (např. pachovou), která by umožnila dalším včelám se orientovat. Pokud si tím nejsme jisti, musíme terče vyměňovat před vpuštěním každé další včely. Dále je třeba zajistit, aby včely nemohli preferovat určitý terč nikoliv kvůli barvě, ale kvůli pozici v pokusném prostoru. Proto bychom ve správně prováděném pokusu pozice barevných terčů náhodně střídali po každé jednotlivé včele. 2. **Před pokusem jsme měli pevně daný celkový rozsah výběru.** Nesprávný (a často užívaný) postup je takový, kdy po prvních 100 včelách zjistíme, zda je výsledek testu průkazný; pokud není, „zvětšíme velikost výběru“, přidáme dalších 30 včel a proceduru opakujeme, a tak to zkusíme několikrát a sledujeme, zda dostaneme kýžený průkazný výsledek, který nám umožní publikovat zásadní práci o tom, jak se včely orientují podle barvy. Takovýto postup mnohonásobně zvyšuje pravděpodobnost chyby prvního druhu!!

(2.) Porovnání poměru pohlaví (*sex ratio*) ve skupině s očekávaným poměrem 1:1. Data: Za poslední měsíc se v porodnici města X narodilo 89 chlapců a 99 děvčat. Byl poměr průkazně odlišný od očekávaného 1:1? Nulová hypotéza tedy zní: Pravděpodobnost narození chlapce a děvčete byla stejná. Očekávané četnosti jsou tedy 94:94.

Ze statistického hlediska nám nic nebrání testovat nulovou hypotézu, že pravděpodobnost narození chlapce byla dvakrát větší než pravděpodobnost narození děvčete. Potom by ovšem očekávané četnosti byly 125.34 a 62.66. Tuto hypotézu bychom jistě zamítli. Ovšem zamítnutí takové hypotézy je málo zajímavé, neboť není žádný důvod předpokládat, proč by měla platit. Naproti tomu, pokud bychom zamítli nulovou hypotézu o poměru 1:1, můžeme hledat smysluplná vysvětlení, proč tomu tak je. Obdobně, při testování štěpných v poměrů v genetice nám ze statistického hlediska nic nebrání testovat nulovou hypotézu, že štěpný poměr je např. 1:17. (Její zamítnutí nám ovšem potvrdí to, co jsme předem věděli, že štěpný poměr 1:17 je zřejmý nesmysl.) Poměr 3:1 očekáváme, protože známe zákony mendelovské dědičnosti, a zároveň víme, že existují mechanismy, které tento poměr narušují. Ty budeme hledat v systémech, kde jsme dokázali odchylku od očekávaného štěpného poměru. (Případné zamítnutí nulové hypotézy o poměru 3:1 tedy nebudeme považovat za

argument pro lysenkovskou biologii.) Formulace nulové hypotézy je tedy, stejně jako postavení pokusu či plán sledování, věcí nejen statistickou, ale především věcí znalosti problému a jeho biologické podstaty.

Nulovou hypotézu formulujeme v matematických termínech, v uvedených příkladech používáme pravděpodobnosti nebo očekávané četnosti jevů. Vše ostatní je mimostatistické uvažování. V příkladu s barevnými tečci jistě mohu usoudit, že pokud zamítну nulovou hypotézu o stejné pravděpodobnosti usednutí na terč, nezávisující na barvě terče, potom včely musí mít schopnost barvy rozlišit. Nicméně hypotéza: včely nejsou schopny rozlišit barvy není nulovou hypotézou statistického testu.

## Velikost výběru

Tento test je pouze přibližný. Přiblížení je velmi dobré, pokud je velikost výběru velká; doporučuje se, aby žádná očekávaná četnost nebyla menší než 1 a aby méně než 20% četností bylo menších než 5. Pokud tomu tak není, můžeme některé kategorie s malými četnostmi spojit.

### Co jsou kritické hodnoty; dosažená hladina významnosti

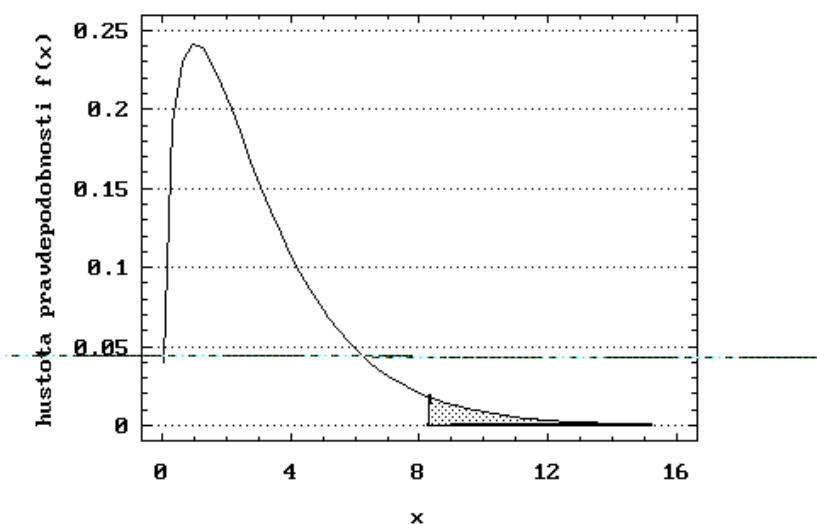
Základem mnoha statistických test je následující postup. Spočteme **testovou statistiku**, např.  $\chi^2$ , o které víme, jaké má rozdělení v případě platnosti nulové hypotézy. Např. víme, že testová statistika, která vznikne sečtením

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - \hat{f}_i)^2}{\hat{f}_i} = \sum_{i=1}^k \frac{(O - E)^2}{E}$$

**Vz. Kapitola 2 -B** O je pozorovaná četnost (*observed*), E je očekávaná četnost (*expected*).

má za předpokladu platnosti nulové hypotézy, na jejímž základě jsou očekávané četnosti počítány, rozdělení, které jsme schopni charakterizovat distribuční funkcí a tuto distribuční funkci vyčíslit. Toto rozdělení se nazývá, stejně jako testová statistika,  $\chi^2$ . Toto rozdělení je spojité; testové kritérium počítáme z počtu případů, tedy nutně z dat diskretních a proto může testové kritérium nabývat pouze určitých hodnot. Proto musí být velikost výběru dostatečně velká, aby „nespojitosť“ příliš nevadila. Toto rozdělení patří mezi tzv. výběrová rozdělení a tvar jeho distribuční funkce závisí na počtu stupňů volnosti. Protože toto rozdělení je známé, lze spočítat jeho 95%-ní kvantil. Víme, že podle definice je 95%-ní kvantil hodnota, kterou náhodná proměnná překročí s pravděpodobností 0.05. Pro shora uvedený test je tedy 95%-ní kvantil rozdělení kritickou hodnotou na 5%-ní hladině významnosti (tj. při  $\alpha = 0.05$ ). Víme, že hodnota kritéria je tím větší, čím je větší odchylka od nulové hypotézy. Pokud tedy hodnota testového kritéria překročí kritickou hladinu na 5%-ní hladině významnosti, můžeme říci, že pokud nulová hypotéza platí, potom pravděpodobnost, že dostaneme výsledek takto nebo více odlišný od nulové hypotézy je menší než 5%. Dnes většina statistických programů přímo s hodnotou testové statistiky poskytuje také odpovídající hodnotu, kterou nejčastěji nazývá *Probability*, případně jenom *P*, ale také někdy *Significance level*. Je to 1 - hodnota distribuční funkce pro spočtenou hodnotu testového kritéria, což je totéž jak hodnota určitého

integrálu z hustoty pravděpodobnosti od spočtené hodnoty do  $+\infty$ . Na grafu hustoty pravděpodobnosti (**Chyba! Nenalezen zdroj odkazů.** obr. kapitola 2 -a) je to plocha, kterou pokrývá „ocas“ rozdělení od dané hodnoty do nekonečna (proto se jim také říká *Tail Area Probabilities*). Tato hodnota nám udává přímo pravděpodobnost, s jakou dostaneme takto nebo více od nulové hypotézy odlišný výsledek za předpokladu, že nulová hypotéza platí. Této hodnotě se říká dosažená hladina významnosti. Pokud je dosažená hladina významnosti menší než 0.05, znamená to, že test je průkazný při  $\alpha = 0.05$ . V biologických člancích se nyní nejčastěji referuje o výsledcích testů následujícím způsobem (výsledek testu z **Chyba! Nenalezen zdroj odkazů.**): V pokusu získaný štěpný poměr 152:39:53:6 se statisticky významně lišil od poměru předpokládaného jednoduchou mendelovskou dědičností ( $\chi^2 = 8.97$ ,  $df=3$ ,  $P<0.05$ ). Pro prezentaci vlastních dat je vhodné dodržovat následující:  $\alpha$  používáme pro předem stanovenou hladinu významnosti, takže píšeme např. „test je průkazný při  $\alpha=0.05$ “;  $P$  (nebo  $p$ ) používáme pro dosaženou hladinu významnosti, takže píšeme  $P<0.05$ . Pokud napíšeme  $P<<0.01$ , znamená to, že dosažená hladina významnosti je výrazně nižší než 0.01; pravděpodobnost chyby prvního druhu je tedy zanedbatelně malá. Pokud nám počítač napíše, že  $P=0$ , znamená to, že hodnota dosažené hladiny významnosti je menší než přesnost počítače. Nepište do článků  $P=0$ , ale např.  $P<10^{-6}$ . Někdy udáváme přímo dosaženou hladinu významnosti. Sdělením  $P=0.49$  naznačujeme, že výsledek testu byl sice průkazný na 5% hladině významnosti, ale „s odřenýma ušima“. Podobně sdělení  $P=0.52$  naznačuje, že jsme nulovou hypotézu sice nezamítli, ale mnoho nechybělo, aby ...Uvádění dosažené hladiny významnosti v publikacích považuji za velmi cennou informaci.



**Obr. Kapitola 2 -A** Hustota pravděpodobnosti rozdělení  $\chi^2$  se dvěma stupni volnosti. Celá plocha vymezená křivkou a osou  $x$  je rovna jedné, velikost tečkované plochy odpovídá pravděpodobnosti, že proměnná nabude hodnotu větší než 8.2. Jestliže jsme dostali hodnotu testového kritéria 8.2, potom velikost tečkované plochy odpovídá dosažené hladině významnosti testu.

Klasická statistika doporučovala striktně postup, kdy nejprve pevně stanovíme hladinu významnosti, a poté dostaneme jednoznačnou odpověď: zamítáme nebo nezamítáme. Dnes se, zvláště v biologické praxi, prosazuje spíše přístup, kdy prezentujeme dosaženou hladinu významnosti a podle ní posuzujeme i důvěryhodnost výsledku: pokud posuzujeme poměr pohlaví v populaci netopýrů, s určitou nejistotou

se smíříme, není třeba zcela jednoznačně rozhodnout ano nebo ne. Naproti tomu prvý způsob musíme nutně použít tam, kde na základě testu činíme rozhodnutí typu ano/ne. Například se rozhoduji, zda zavést výrobu preparátu, který v populaci mění poměr pohlaví. Zde si musím předem stanovit míru rizika, kterou jako výrobce hodlám nést (bude pravděpodobně velmi malá, například 0.001), a pokud výsledek pokusu nebude průkazný, výrobu nezavedu. Obdobný postup se užívá u klinických testů při zavádění nových léků.

### **Příliš dobré, aby to byla pravda (*Too good to be true*)**

Příklad: Společnost, vyrábějící nový druh žvýkačky byla obviněna, že pravidelné žvýkání jejích produktů vede u mužů ke zvýšené mortalitě spermií nesoucích chromosom X (a pak se jim budou rodit převážně synové). Společnost najala pokusnou osobu, která dva roky intenzivně žvýkala její produkty; poté provedla vyšetření jeho spermatu. Ve zveřejněné zprávě uvádí, že provedla test na přítomnost chromosomu X v 10000 spermií pokusné osoby, a zjistila přítomnost chromosomu X v 5001 případech, tzn. nepřítomnost v 4999 případech. Společnost na základě toho konstatuje, že shoda s očekávaným poměrem 1:1 je dostatečně jasná a že tedy její produkty jsou z tohoto hlediska zcela neškodné. Co k tomu můžeme říci jako statistici?

Odhlédněme nyní od toho, že pozorování nebylo nejlépe naplánováno (eufemismus, říkající, že bylo naplánováno úplně špatně), chybí kontrola, jedná o vliv na jediného člověka, a pokusme se vyhodnotit porovnání poměru spermií s chromosomem k počtu spermií s chromosomem Y s očekávaným poměrem 1:1. Použijme  $\chi^2$ - test dobré shody. Dostáváme  $\chi^2 = 4.10^{-4}$ ,  $P=0.984$ . Výsledek testu je tedy neprůkazný, ale dosažená hladina významnosti se blíží jedné. To je velmi podezřelé. Co nám to říká? Předpokládejme, že poměr 1:1 je v základním souboru opravdu zachován. Potom s pravděpodobností více než 98% dostaneme v náhodném výběru spermií větší odchylku od poměru 1:1, než v našich datech. Nebo jinak: šance, že dostaneme takto dobrou shodu s poměrem 1:1 byla menší než 2%. Buď tedy měla společnost z pekla štěstí, nebo spíše výsledky zfalšovala tak, jak ji vyhovovaly. Jsou příliš dobré na to, aby to mohla být pravda.

Uvedený příklad je jistě vymyšlený. Nicméně ukazuje na to, jak se dá statistikou objevit falšování dat. Obdobný (statisticky) příklad je ovšem ve světové vědě znám: Vyhodnotíme-li uvedeným postupem výsledky originálních Mendelových pokusů, zjistíme, že jsou z uvedeného hlediska „příliš dobré“, shoda se štepnými poměry je nepravděpodobně dobrá. Mendel sám ovšem o statistice netušil a nikde netvrdí, že se řídil pravidly pro náhodný výběr; naopak, konstatuje, že tam, kde byl výběr malý přidával další individua. Závěr, hojně citovaný v době, kdy u nás byla genetika nazývána buržoasní pavědou, že „prelát Mendel falšoval data“, je tedy nesmyslný; pro zájemce doporučuji článek T. Havránka (1986). Poučení pro nás je ovšem dvojí: když přinášíme zprávu o výsledku pokusu, popišme detailně, jak jsme k datům přišli a při použití historických dat nemůžeme předpokládat, že data byla sebrána způsobem odpovídajícím statistickým zásadám.



## Doporučená četba

Zar (1984), pp. 40-60, Sokal-Rohlf (1981), pp. 692-730, Havránek (1993): testování hypotéz obecně - 73-95.

Havránek T. (1986): Gregor Mendel a experimentální data. - Vesmír 65: 331-333.

**Tabulka Kapitola 2 -C** tabulka kritických hodnot  $\chi^2$  distribuce

df	$\alpha$		
	0.05	0.01	0.001
1	3.841	6.635	10.827
2	5.991	9.210	13.815
3	7.815	11.345	16.268
4	9.488	13.277	18.465
5	11.070	15.086	20.517
6	12.592	16.812	22.457
7	14.067	18.475	24.322
8	15.507	20.090	26.125
9	16.919	21.666	27.877
10	18.307	23.209	29.588
11	19.675	24.725	31.264
12	21.026	26.217	32.909
13	22.362	27.688	34.528
14	23.685	29.141	36.123
15	24.996	30.578	37.697
16	26.296	32.000	39.252
17	27.587	33.409	40.790
18	28.869	34.805	42.312
19	30.144	36.191	43.820
20	31.410	37.566	