

STATISTICKÉ
ZPRACOVÁNÍ DAT –
FISHERŮV EXAKTNÍ TEST

Metoda chí kvadrát x Fisherův test

- Pro zjištění závislosti – metoda chí kvadrát
- V některých případech metodu chí kvadrát nelze použít
 - rozsah souboru menší než 20
 - očekávané četnosti jsou malé
- Lze použít Fisherův test - založen na jiném principu
- Fisherův exaktní test je založen na výpočtu přesné (exaktní) pravděpodobnosti, se kterou bychom za platnosti nulové hypotézy o nezávislosti veličin získali naši konkrétní realizaci kontingenční tabulky

Fisherův test

- Zjišťujeme závislost dvou kvalitativních veličin na prvcích téhož výběru
- Máme náhodný výběr rozsahu n rozdělený do dvou skupin (skupina 1, skupina 2)
- Skupiny mohou nabývat hodnotu jednoho ze dvou znaků (znak 1, znak 2)
- Příkladem - skupina ženy, muži, znak kouří, nekouří
- Úkolem testu je rozhodnout, zda znaky jsou na sobě závislé nebo nezávislé (zda znak 1 má vliv na znak 2)
- Fisherův exaktní test odvozen pro kontingenční tabulku 2x2 tzv. čtyřpolní tabulku, ale existuje i jeho zobecnění pro libovolnou kontingenční tabulku

ZÁKLADNÍ PRINCIP FISHEROVA TESTU

- Testujeme nulovou hypotézu proti alternativní hypotéze.
- Nulová hypotéza H_0 : znaky 1 a 2 jsou nezávislé (Pozorované četnosti by měly odpovídat očekávaným četnostem)
- Alternativní hypotéza H_1 : Mezi znaky 1, 2 je závislost
- Nepředpokládá se, že teoretické rozdělení četností je známé, ale počítá se přímo pravděpodobnost odchylky od nulové hypotézy
- Při testování se generují varianty pozorované tabulky četností a určuje se pravděpodobnost výskytu všech obměn, které mají stejné součty okrajových četností
- Hlavní myšlenkou testu je výpočet pravděpodobnosti, se kterou bychom získali čtyřpolní tabulky stejně nebo více vzdálené od nulové hypotézy při zachování marginálních četností

VÝPOČET TESTOVÉ STATISTIKY

	Znak 1	Znak 2	Součet
Skupina 1	a	b	$a+b$
Skupina 2	c	d	$c+d$
Součet	$a+c$	$b+d$	n

Čtyřpolní tabulka

a, b, c, d četnosti
 $a+b, c+d, a+c, b+d$ okrajové četnosti tzv. marginální četnosti.

- Z hodnot a, b, c, d se vybere hodnota a a od té se postupně odečítá a po té přičítá hodnota 1, aby součet okrajových četností zůstal stejný a byly vyčerpány všechny možné případy. Např. pokud se od hodnoty a odečte 1, musí se k hodnotě b přičíst 1, k hodnotě c přičíst 1 a od hodnoty d odečíst 1, aby okrajové četnosti zůstaly stejné
- Generují se všechny možné varianty tabulky četností
- Pro původní a každou vygenerovanou tabulku se vypočítá pravděpodobnost

Vzorec pro výpočet pravděpodobnosti

$$p_i = \frac{\binom{a+c}{a} \binom{b+d}{b}}{\binom{n}{a+b}} = \frac{(a+b)! (c+d)! (a+c)! (b+d)!}{n! a! b! c! d!}$$

p_i pravděpodobnost vypočtená z tabulky i

a, b, c, d četnosti uvnitř tabulky i

n rozsah souboru

Hodnota testového kritéria

- Hodnota testového kritéria testové statistiky je součet všech vypočtených pravděpodobností menších nebo stejných jako hodnota pravděpodobnosti, která přísluší čtyřpolní tabulce sestrojené na základě pozorovaných hodnot
- Hodnota testového kritéria se porovnává s hladinou významnosti α
- V případě oboustranného testu se sčítají hodnoty všech vypočtených pravděpodobností u tabulek, které jsou menší nebo rovny než skutečná zjištěná četnost
- Pokud je $\sum p_i < \alpha$, potom nulovou hypotézu o nezávislosti zamítáme a přijímáme hypotézu, že určitá závislost existuje

Příklad 1

- Skupina 1 a 2, znak 1 a 2, zkoumáme závislost mezi skupinami a znaky
- Hladina významnosti 5 %
- Ze získaných dat vytvoříme čtyřpolní tabulku

	Znak 1	Znak 2	Součet
Skupina 1	2	5	7
Znak 2	3	2	5
Součet	5	7	12

- Z této tabulky vybereme hodnotu 2 (skupina 1, znak 1) a od hodnoty 2 postupně odečítáme 1 a po té přičítáme hodnotu 1
- Ostatní hodnoty doplňujeme tak, aby součet okrajových četností zůstal stejný
- Dostaneme následující tabulky:

	Znak 1	Znak 2	Součet
Skupina 1	0	7	7
Znak 2	5	0	5
Součet	5	7	12

	Znak 1	Znak 2	Součet
Skupina 1	1	6	7
Znak 2	4	1	5
Součet	5	7	12

	Znak 1	Znak 2	Součet
Skupina 1	3	4	7
Znak 2	2	3	5
Součet	5	7	12

	Znak 1	Znak 2	Součet
Skupina 1	4	3	7
Znak 2	1	4	5
Součet	5	7	12

	Znak 1	Znak 2	Součet
Skupina 1	5	2	7
Znak 2	0	5	5
Součet	5	7	12

	Znak 1	Znak 2	Součet
Skupina 1	2	5	7
Znak 2	3	2	5
Součet	5	7	12

$$\begin{aligned}
 p &= \frac{\binom{a+c}{a} \binom{b+d}{b}}{\binom{n}{a+b}} = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!} = \frac{(2+5)!(3+2)!(2+3)!(5+2)!}{12!2!5!3!2!} = \frac{7!5!5!7!}{12!2!5!3!2!} \\
 &= \frac{7!5!5!7!}{12!2!5!3!2!} = \frac{7!5!5!7!}{7! \cdot 8 \cdot 9 \cdot 10 \cdot 11 \cdot 12 \cdot 2! \cdot 5! \cdot 3! \cdot 2!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2}{8 \cdot 9 \cdot 10 \cdot 11 \cdot 12 \cdot 2 \cdot 3 \cdot 2 \cdot 2} = \frac{7 \cdot 6 \cdot 5}{8 \cdot 9 \cdot 11} = \frac{7 \cdot 3 \cdot 5}{4 \cdot 9 \cdot 11} = \frac{7 \cdot 5}{4 \cdot 3 \cdot 11} = \frac{35}{132}
 \end{aligned}$$

= 0,265152

Vložit funkci ? ×

Vyhledat funkci:

Přejít

Vybrat kategorii: Naposledy použité

Vybrat funkci:

FAKTORIÁL

KDYŽ

CHISQ.TEST

CHISQ.INV

QUARTILEX

SMODCH.VÝBĚR.S

MAX

FAKTORIÁL(číslo)
 Vrátí faktoriál čísla. Výsledek se rovná hodnotě 1*2*3*...*Číslo.

- Pravděpodobnost pro čtyřpolní tabulku sestavenou na základě pozorovaných hodnot je 0,265152
- Menší nebo stejné hodnoty nabývají pravděpodobnosti p_1 , p_2 , p_3 , p_5
- Hodnota testové statistiky (p -hodnota) je součet všech vypočtených pravděpodobností menších nebo stejných jako hodnota pravděpodobnosti pro čtyřpolní tabulku sestavenou na základě pozorovaných hodnot, tzn., že je

$$\sum p_i = p_1 + p_2 + p + p_3 + p_5 = 0,001263 + 0,044192 + 0,265152 + 0,220960 + 0,026515 = 0,558082$$
- Není splněna podmínka $\sum p_i < \alpha$, platí $\sum p_i > 0,05$, nulovou hypotézu o nezávislosti přijímáme a lze konstatovat, že mezi skupinami 1 a 2 a znaky 1 a 2 není závislost

p -hodnota

- p -hodnota je nejmenší hladina významnosti, při které ještě zamítneme nulovou hypotézu
- p -hodnota je pravděpodobnost, že při platnosti nulové hypotézy nabývá testová statistika své stávající hodnoty anebo hodnot ještě extrémnějších (nepříznivějších vůči nulové hypotéze)
- p -hodnota je pravděpodobnost, s jakou bychom mohli obdržet pozorovaná data nebo data stejné, či ještě více odporující nulové hypotéze, za předpokladu, že je nulová hypotéza pravdivá. Čím menší je p , tím neudržitelnější čili méně důvěryhodná je nulová hypotéza

- Generování všech možných variant tabulky četností je poměrně pracné, ale existuje řada programů, kde stačí zadat hodnoty zjištěných četností do tabulky a výsledkem je hodnota testové statistiky
- Příkladem vhodného programu je odkaz na <http://www.langsrud.com/fisher.htm>
- Aplikaci, která umožňuje zobecnění na kontingenční tabulku max 2x5 <https://quantitativeskills.com/sisa/statistics/fiveby2.htm>



https://www.langsrud.com/fisher.htm

[Domů](#) | [50-50 MANOVA](#) | [Rotační testy](#) | [Software](#) | Fisherův přesný test | [Publikace](#)

Fisherův přesný test

VYPOČÍTAT

3	1
1	3

PŘEHLEDNÝ STŮL

ČISTÝ VÝSTUP

[ÚVOD](#)

Vytvořil

©

[Øyvind Langsrud](#)

[Německá verze](#)

Fisher's Exact Test
https://www.langsrud.com/fisher.htm

ÚVOD

Na této stránce lze statisticky otestovat, zda existuje nějaký vztah mezi dvěma kategorickými proměnnými (se dvěma úrovněmi). Vyplňte tabulku a stiskněte COMPUTE.

Výstup se skládá ze tří p-hodnot:

- **Vlevo:** Použijte tuto možnost, pokud je alternativou k nezávislosti to, že existuje negativní Asociace mezi proměnnými. To znamená, že pozorování mají tendenci ležet vlevo dole a vpravo nahoře.
- **Vpravo:** Použijte tuto možnost, pokud je alternativou k nezávislosti to, že existuje pozitivní Asociace mezi proměnnými. To znamená, že pozorování mají tendenci ležet vlevo nahoře a vpravo dole.
- **2-ocas:** Tuto možnost použijte, pokud neexistuje žádná předchozí alternativa.

POZNÁMKA: Rozhodněte se použít levý, pravý nebo 2-Tail předtím shromažďování (nebo prohlížení) dat.

Příkladem takových údajů (neskutečných) je 100 osob klasifikovaných po sexu a e-mailu adresa:

	ženský	muž
E-mailová adresa	3	15
bez e-mailové adresy	37	45

tj. 3 ze 40 žen mají e-mailovou adresu a 15 z 60 mužů má e-mailovou adresu.

Údaje mohly být shromažďovány různými způsoby:

1. Zeptali jsme se 60 mužů a 40 žen. To znamená, že celkový počet mužů a žen je pevně daný.
2. Zeptali jsme se 100 osob na jejich pohlaví a na to, zda mají e-mail adresu. To znamená, že předem je stanoven pouze celkový počet osob.
3. Zeptali jsme se všech osob v Norsku narozených 24/11-68. Tj.: Celkový počet osob není předem stanovena.

Dvoustranná p-hodnota se vypočítá podle definice v Agresti (1992) § 2.1. b).

ODKAZ: Agresti A, (1992), Přehled exaktní inference pro kontingenční tabulky, *Statistical Science*, 7,131-153

	Znak 1	Znak 2	Součet
Skupina 1	2	5	7
Znak 2	3	2	5
Součet	5	7	12

Fisherův přesný test

VYPOČÍTAT

2	5
3	2

PŘEHLEDNÝ STŮL

ČISTÝ VÝSTUP

[ÚVOD](#)

Vytvořil

©

[Oyvind Langsrud](#)

[Německá verze](#)

Fisher's Exact Test
<https://www.langsrud.com/fisher.htm>

 TABLE = [2 , 5 , 3 , 2]
 Left : p-value = 0.31060606060606255
 Right : p-value = 0.954545454545455
 2-Tail : p-value = 0.5580808080808117

$$\sum p_i = p_1 + p_2 + p_3 + p_4 + p_5 =$$

$$= 0,001263 + 0,044192 + 0,265152 + 0,220960 + 0,026515 =$$

$$= 0,558082$$



qs Fisherův přesný test pro křížový stůl 2*5 nebo r

https://quantitatives... Aktualizovat (Ctrl+R)

SISA

Přesně pět krát dva

R1: 2;
5 ks

R2: 3;
2 ks

R3: 0;
0

R4: 0;
0

R5: 0;
0

R6: 0;
0

R7: 0;
0

bodprob: 0.26515
pro získání přesně této tabulky vzhledem k margináliímchí-kvadrát
oba s 1 stupněm volnosti
Pearson: 1.185 (p= 0.27628)
LRX: 1.195 (p= 0.27437)

Více:

Analýza

tabulky dva po dvou chí-kvadráty, poměry rizika, poměry šancí
a další statistiky tabulky T-test, NNT, PARF

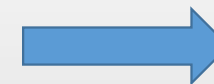
první sloupec proti součtům řádků

první řádek proti součtům

sloupců

[Nápovědu najdete na webu SISA](#)

Tabulka 2x2



https://quantitatives... a A ☆ □ ☆

SISA

Tabulková analýza dva krát dva

Řádek 1: 2 5
Řádek 2: 3 2
CI: 95 %
z pro 95% CI= 1.96

Celkový počet případů v tabulce 12
Nejmenší buňka obsahuje 2 případy
Zvažte použití [Fisher](#).

Chi kvadráty
(všechny s 1 stupněm volnosti): Pearson's= 1.185 (p=0.276278) *
Věrohodnostní poměr= 1.195 (p=0.27437)
Yate's= 0.245 (p=0.62069)
Mantel Haenszel= 1.087 (p=0.29724)

[Nápovědu najdete na webu SISA](#)

Více:

[Fisherův exaktní test](#) [T-test](#), [NNT](#), [PARF](#)
[první sloupec proti součtům řádků](#)
[první řádek proti součtům](#)
sloupců

Fisherův exaktní test





SISA

Fisherův exaktní test

Řádek 1: 2; 5

Řádek 2: 3; 2

Celkový počet případů= 12

Nejmenší hodnota= 2

Nejmenší mezní= 5

Doporučuje se používat statistiky * (označené hvězdičkou)

p přesně pro tuto tabulku= 0.26515

Jednostranné p-hodnoty: pro $p(O \geq E)$:

$p(O \geq E)$: 0,3106061 *

$p(O > E)$: 0,0454545

střední p: 0,1780303

$p(O \leq E)$: 0,9545455

Oboustranné p-hodnoty $p(O \geq E | O \leq E)$:

$p = 0.5580808$ * (součet malých p)

$p = 0.2651515$ (vlevo+vpravo-přesně)

$p = 0.3560606$ (vlevo+vpravo)

$p = 0.621212$ (dvojnásobek jednostranného p)

$p(O > E | O < E) = 0.29293$ (součet malých p's)

mid-p 0.4255051 (součet malých p's)

[Nápovědu najdete na webu SISA](#)

Výpočet byl

$p = 0,558082$

Příklad 2

- <http://portal.matematickabiologie.cz/index.php?pg=aplikovana-analyza-klinickyh-a-biologickyh-dat--analyza-a-management-dat-pro-zdravotnicke-obory--testovani-hypotez-o-kvalitativnich-promennych--fisheruv-exaktni-test>

The screenshot shows a web browser window with the URL <https://portal.matematickabiologie.cz/index.php?pg=aplikovana-analyza-klinickyh-a-biologickyh-dat--analyz...>. The page is titled "Aplikovaná analýza klinických a biologických dat" and contains a sidebar with navigation links and a main content area. The main content area is titled "Fisherův exaktní test" and includes a definition of a 2x2 contingency table, a table labeled "Tabulka 4: Ukázka čtyřpolní tabulky," and a paragraph explaining the test's application in medicine and biology.

Fisherův exaktní test

Definice čtyřpolní tabulky je zřejmá - je to nejjednodušší možná kontingenční tabulka, kdy obě sledované náhodné veličiny mají pouze dvě varianty, kterých mohou nabývat. Stejně jako v případě obecné kontingenční tabulky můžeme pomocí statistických metod rozhodovat o statistické závislosti dvou sledovaných veličin, v případě čtyřpolní tabulky můžeme navíc velmi jednoduše rozhodovat i o míře této závislosti (o těsnosti statistické vazby). Příklad čtyřpolní tabulky představuje tabulka 4, kde jsou četnosti jednotlivých možných kombinací náhodných veličin X a Y označeny písmeny a, b, c a d .

Tabulka 4: Ukázka čtyřpolní tabulky.

Náhodná veličina X	Náhodná veličina Y		Celkem
	$Y=1$	$Y=2$	
$X=1$	a	b	$a+b$
$X=2$	c	d	$c+d$
Celkem	$a+c$	$b+d$	$a+b+c+d$

Při rozhodování o nezávislosti ve čtyřpolní tabulce můžeme samozřejmě použít Pearsonův chí-kvadrát test, neboť tento test lze použít na jakoukoliv kontingenční tabulku, nicméně u tohoto testu je nutné hlídat jeho předpoklady: 80 % očekávaných četností, e_{ij} , větších než 5 totiž v případě čtyřpolní tabulky znamená 100 % očekávaných četností, které mají být větší než 5. Nedodržení předpokladů pro Pearsonův chí-kvadrát test může stejně jako u t -testu a analýzy rozptylu vést k nesmyslným závěrům. Situace s malými pozorovanými a tedy i očekávanými četnostmi jsou ale bohužel v medicíně i biologii relativně časté, a to samé platí i pro čtyřpolní tabulky. Zlatým standardem pro hodnocení čtyřpolních tabulek se proto stal jiný test, tzv. Fisherův exaktní test (*Fisher exact test*), který je založen na výpočtu přesné (exaktní) pravděpodobnosti, se kterou bychom za platnosti nulové hypotézy o nezávislosti veličin X a Y získali naši konkrétní realizaci čtyřpolní tabulky.