

Testovanie štatistických hypotéz I.

Pravdepodobnosť a štatistika pre informatikov 2025

Príklad 1

Ernest Rutherford spolu so svojimi spolupracovníkmi skúmal vo svojom známom experimente rádioaktívne častice, ktoré dopadali na detektor. Počas 2608 časových intervalov dĺžky 8 sekúnd sa zaznamenával počet dopadnutých častíc.

Vektor **rutherford** obsahuje počet časových intervalov, počas ktorých bol zaznamenaný počet častíc 0, 1, 2, ..., 14.

```
rutherford = [57, 203, 383, 525, 532, 408, 273, 139, 45, 27, 10, 4, 0, 1, 1]
```

Testujte hypotézu, že počet častíc, ktoré zaznamenal detektor má Poissonovo rozdelenie.

Príklad 2

Testujte hypotézu o Poissonovom rozdelení pre dáta, ktoré vyjadrujú počet oznámení o úmrtiach v londýnskych novinách Times v jednotlivých dňoch počas troch rokov (dáta boli publikované v roku 1914). V prvom prípade ide o mužov vo veku nad 85 rokov, druhom prípade o ženy vo veku nad 80 rokov.

```
times_1 = [484, 391, 164, 45, 11, 1]
```

```
times_2 = [162, 267, 271, 185, 111, 61, 27, 8, 3, 1]
```

V oboch prípadoch ide o početnosti hodnôt 0, 1, 2, atď.

Príklad 3

Testujte hypotézu o Poissonovom rozdelení pre dáta, ktoré vyjadrujú počet prípadov, kedy bol volajúci prepojený na nesprávne telefónne číslo (hodnoty sú približne spred 100 rokov) v sledovanom časovom úseku.

```
telefon = [1, 5, 11, 14, 22, 43, 31, 40, 35, 20, 18, 12, 7, 6, 2]
```

V tomto prípade početnosti zodpovedajú 2, 3, 4,... nesprávnym prepojeniam.

Príklad 4

V roku 1894 spravil štatistik Walter F. R. Weldon experiment, v ktorom opakovane hádzal 12 kockami (spolu 26 306 krát) a zaznamenával, na koľkých kockách padla päťka alebo šesťka. Počet týchto kociek v jednotlivých pokusoch (0, 1, 2, ..., 12) bol

weldon = [185, 1149, 3265, 5475, 6114, 5194, 3067, 1331, 403, 105, 14, 4, 0]

Niektorí už v bonusovej úlohe zistili, že nejde o binomické rozdelenie $\text{Bin}(12, \frac{1}{3})$. Z princípu experimentu však vyplýva, že by malo ísť o *nejaké* binomické rozdelenie. Odhadnite jeho parameter a otestujte zhodu dát s týmto binomickým rozdelením.

Príklad 5

Nasledujúce hodnoty sú časy čakania v sekundách zákazníkov vo fast-foode doobeda (AM) a poobede (PM).

<https://cran.r-project.org/web/packages/ipsRdbs/refman/ipsRdbs.html#ffood>

AM = [38, 100, 64, 43, 63, 59, 107, 52, 86, 77]

PM = [45, 62, 52, 72, 81, 88, 64, 75, 59, 70]

Otestujte zhodu pravdepodobnostného rozdelenia času čakania doobeda a poobede.

Príklad 6

Náhodné premenné X , Y sú nezávislé a každá z nich má exponenciálne rozdelenie s tou istou strednou hodnotou. Zvoľte si niekoľko (2-3) hodnôt pre túto spoločnú strednú hodnotu náhodných premenných. Zistite (t. j. vyslovte na základe simulácií hypotézu), aké rozdelenie má náhodná premenná $(X-Y)/(X+Y)$.

Otestujte svoje tvrdenie pomocou Kolmogorovovho-Smirnovovho testu.

Príklad 7

V bakalárskej práci *Štatistické hypotézy pri modelovaní domácej výhody vo futbalových zápasoch* (Matúš Košút, 2024) boli analyzované aj tieto dáta z Premier League:

	Karty hostí j			
Karty domácich i	0	1	2	3+
0	22	24	18	10
1	20	41	31	24
2	11	28	23	36
3+	8	26	22	36

Tabuľka 18: Kontingenčná tabuľka žltých kariet domácich a hosťujúcich mužstiev v sezóne 2021/22.

Počet žltých kariet vo futbalovom zápase je náhodná premenná. Testujte nezávislosť počtu žltých kariet pre domácich a pre hostí.

Príklad 8

Otestujte zhodu s Berfordovým rozdelením pre niektoré z dát z článku

<https://www.sciencedirect.com/science/article/abs/pii/S0378437114002854>

Príklad 9

Pre dáta (zdroj: <https://rdr.io/github/carloscinelli/benford.analysis/man/sino.forest.html>) zo súboru http://www.iam.fmph.uniba.sk/institute/stehlikova/pas25/cv/sino_forest.csv získajte ich prvú číslicu a otestujte zhodu získaných hodnôt s Benfordovým zákonom.

Príklad 10 (bonus pre 2AIN1 – naprogramovanie tejto hry a štatistické vyhodnotenie výsledkov)

Naprogramujte jedno kolo nasledovnej hry (to znamená, že ignorujeme inštrukciu “he is also the starter of the new game” – vždy sa odohrá len jedna hra). Spravte simulácie hry a zaznamenávajte, ktorý hráč vyhral (prvý, druhý alebo tretí). Testujte hypotézu, že pravdepodobnosti víťazstva všetkých troch hráčov sú rovnaké.

The Mang Kung (“blind men”) Dice Game is an old Chinese game played with six non-traditional dice. These dice, which are referred to as Mang Kung Dice, are marked as follows:

Die 1:	1					
Die 2:		2				
Die 3:			3			
Die 4:				4		
Die 5:					5	
Die 6:						6

where the symbol $\square$ represents BLANK. In other words, five faces of each *Mang Kung Die* are not marked. Only one face of each cube is marked according to the traditional way.

The Mang Kung Dice Game is still a fairly popular game among Chinese gamblers in Southern China and Hong Kong. In the next section we describe the game.

◆ DESCRIPTION OF THE GAME ◆

We first describe the game with three players. Each player contributes 7 units of stake into a “pool”. Thus, the total initial amount in the pool is 21 units. A player is randomly selected to start the first game. Each time six *Mang Kung Dice* are thrown simultaneously. The total face value of these dice is counted. This total face value (F) is then compared with the total amount in the pool (P). The possible outcomes and the corresponding player’s actions are listed in Table 1.

Table 1. The three possible situations in the Mang Kung Dice Game.

Outcomes	Player’s Actions
$F < P$	The Player can take the amount F from the pool. He then passes the dice to the next player on his left to continue the game.
$F > P$	The Player has to top up the pool by the difference, $(F - P)$. He then passes the dice to the next player to continue the game.
$F = P$	The Player wins the game. He receives the amount P from the pool. Moreover, each of the other players has to pay him an amount F as “bonus”. He is also the starter of the next new game.