

I. Náhodné vektory: simulácie

Metódy riešenia úloh z pravdepodobnosti a štatistiky
Beáta Stehlíková, FMFI UK Bratislava

Náhodné vektory s nezávislými zložkami

- Prakticky sme už takúto situáciu mali:

Two insurers provide bids on an insurance policy to a large company. The bids must be between 2000 and 2200. The company decides to accept the lower bid if the two bids differ by 20 or more. Otherwise, the company will consider the two bids further. Assume that the two bids are independent and are both uniformly distributed on the interval from 2000 to 2200. Determine the probability that the company considers the two bids further.

(A) 0.10

(B) 0.19

(C) 0.20

(D) 0.41

(E) 0.60

```
pr1 <- function(){  
  x <- runif(1, min = 2000, max = 2200)  
  y <- runif(1, min = 2000, max = 2200)  
  return(abs(x - y) < 20)  
}
```

Podmienené rozdelenia

- Simulácia je priamočiara, napríklad

An auto insurance policy will pay for damage to both the policyholder's car and the other driver's car in the event that the policyholder is responsible for an accident. The size of the payment for damage to the policyholder's car, X , has a marginal density function of 1 for $0 < x < 1$. Given $X = x$, the size of the payment for damage to the other driver's car, Y , has conditional density of 1 for $x < y < x + 1$. If the policyholder is responsible for an accident, what is the probability that the payment for damage to the other driver's car will be greater than 0.500?

(A) $3/8$

(B) $1/2$

(C) $3/4$

(D) $7/8$

(E) $15/16$

```
pr2 <- function(){  
  x <- runif(1)  
  y <- runif(1, min = x, max = x + 1)  
  return(y < 0.5)  
}
```

Zadaná združená hustota

- Nabudúce odvodíme postup, napríklad:

Let X and Y be random losses with joint density function

$$f(x, y) = e^{-x-y} \quad \text{for } x > 0 \text{ and } y > 0.$$

An insurance policy is written to reimburse $X + Y$. Calculate the probability that the reimbursement is less than 1.

(A) e^{-2}

(B) e^{-1}

(C) $1 - e^{-1}$

(D) $1 - 2e^{-1}$

(E) $1 - 2e^{-2}$

II. Testy dobrej zhody

Testy dobrej zhody

- Chceme overiť, či naše dáta pochádzajú z konkrétneho pravdepodobnostného rozdelenia.
- Napríklad:
 - Je kocka pravidelná? Padajú na nej čísla 1, 2, 3, 4, 5, 6 rovnako často?
 - Je pravdepodobnosť výhry jednotlivých hráčov v simulovaných priebehoch rovnaká?
 - Dostávame simuláciami v domácej úlohe náhodné čísla z rozdelenia, ktoré sme sa snažili uhádnuť?

Príklad: Je kocka pravidelná?

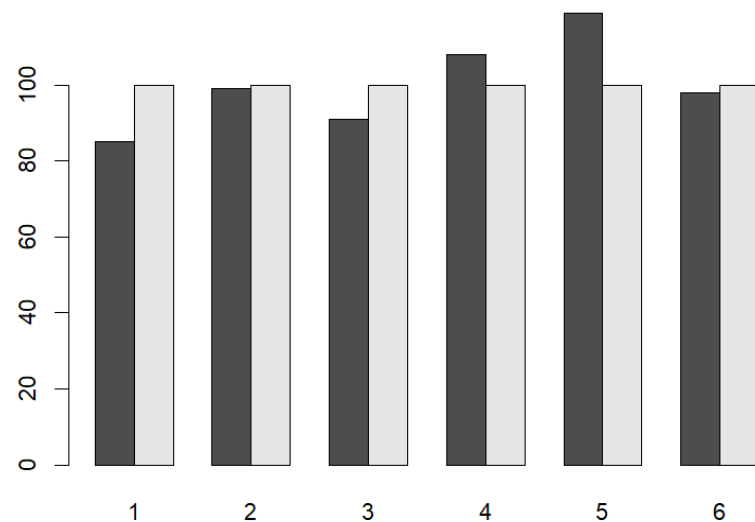
- Môžeme kocku považovať za pravidelnú?
- Výsledky 600 hodov:

padnuté číslo	početnosť
1	85
2	99
3	91
4	108
5	119
6	98

- Očakávané počty sú všetky 100, nakoľko významné sú rozdiely?

- Očekávané počty a skutečné:

```
skutocne <- c(85, 99, 91, 108, 119, 98)  
ocakavane <- rep(100, 6)  
barplot(rbind(skutocne, ocakavane), beside = TRUE, names.arg = 1:6)
```



- Pozrieme sa na rozdiely „skutočné mínus očakávané“
- Na čo treba myslieť:
 - Kladné a záporné odchýlky sa nemôžu navzájom zrušiť
 - Rozdiel -15 nemá rovnakú váhu, keď je očakávaný počet 100 a keď je 1000

```
> tab <- data.frame(skutocne, ocakavane)
> tab$rozdiel <- with(tab, skutocne - ocakavane)
> tab
```

	skutocne	ocakavane	rozdiel
1	85	100	-15
2	99	100	-1
3	91	100	-9
4	108	100	8
5	119	100	19
6	98	100	-2

- Preto:
 - Rozdiel na druhú
 - Relatívne voči očakávaným počtom

```
> tab$rozdiel2rel <- with(tab, rozdiel^2/ocakavane)
> tab
```

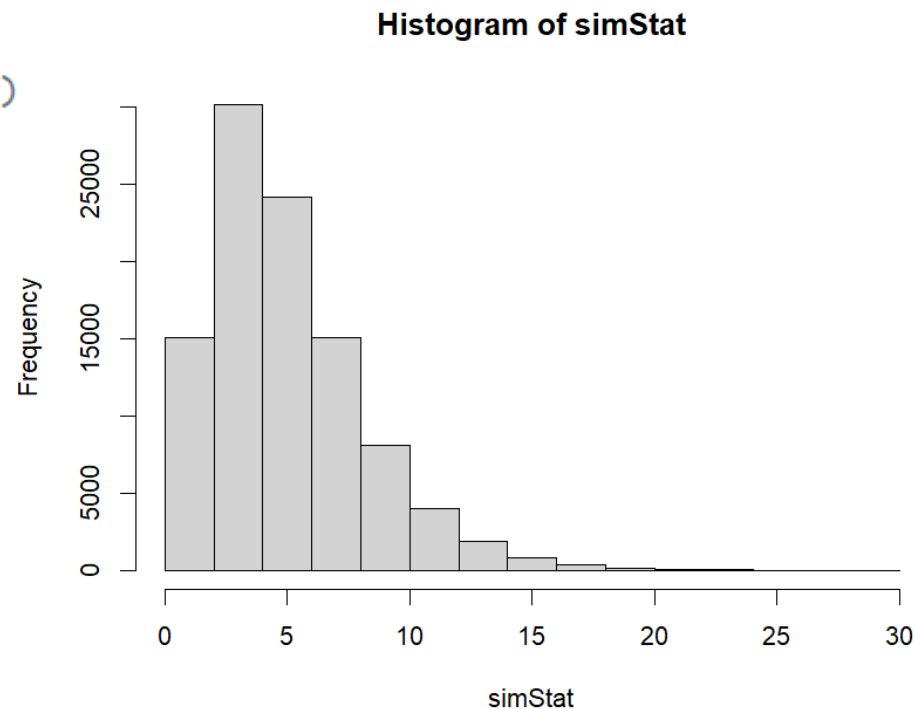
	skutocne	ocakavane	rozdiel	rozdiel2rel
1	85	100	-15	2.25
2	99	100	-1	0.01
3	91	100	-9	0.81
4	108	100	8	0.64
5	119	100	19	3.61
6	98	100	-2	0.04

- Súčet týchto hodnôt bude „mierou rozdielu“, hypotézu o rozdelení zamietneme, ak bude veľký

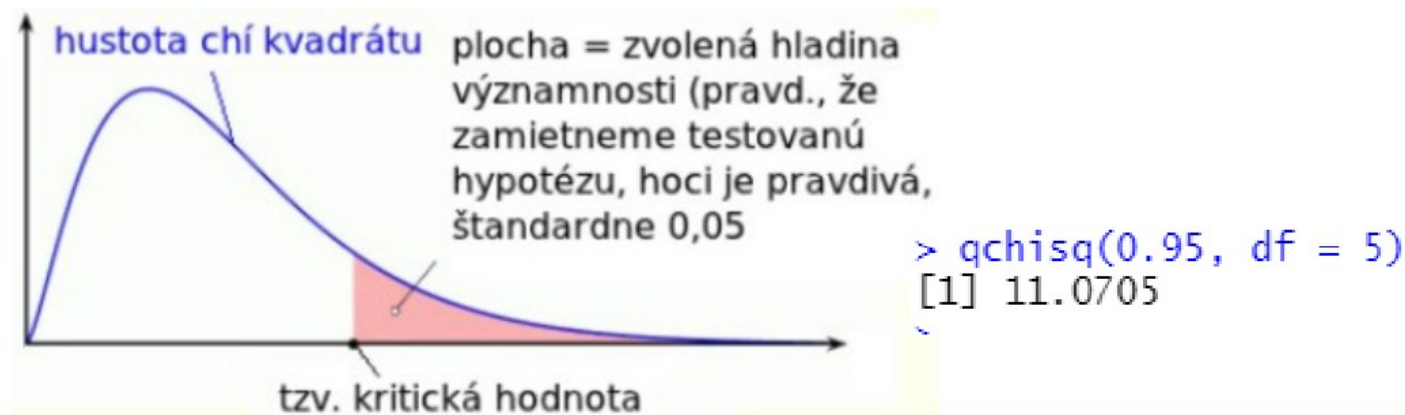
- Čo znamená „veľký rozdiel“:

```
statistika <- function(){  
  kocky <- sample(1:6, size = 600, replace = TRUE)  
  pocty <- table(factor(kocky, levels = 1:6))  
  stat <- sum((pocty - 100)^2/100)  
  return(stat)  
}  
set.seed(123)  
simStat <- replicate(10^5, statistika())
```

```
> quantile(simStat, 0.95)  
95%  
11.06
```



- Čo znamená „veľký rozdiel“:
 - Väčší ako kvantil chí kvadrát rozdelenia s $q-1$ stupňami voľnosti, kde q je počet skupín, do ktorých sú dáta zaradené (teraz je teda $q=6$)
- Chí kvadrát rozdelenie je aproximácie, ktorá sa dá použiť, ak je očakávaný počet dát v každej skupine aspoň 5



- V našom prípade:

```
> tab
  skutocne ocakavane rozdiel rozdiel2rel
1       85        100     -15        2.25
2       99        100      -1         0.01
3       91        100      -9         0.81
4      108        100       8         0.64
5      119        100      19         3.61
6       98        100      -2         0.04
> sum(tab$rozdiel2rel)
[1] 7.36
```

- Pravidelnosť kocky teda nezamietame.