

Matematika 6F
**Rozdělení počtu zpráv poslaných
do mailing listu**

MARTIN BRUCHANOV,

cv. pondělí 14:30, letní semestr 2005/2006

30. května 2006

1. Zadání

Otestovat hypotézu H_0 : rozložení počtu denních zpráv přijatých do emailové konference během měsíce dubna má rovnoměrné rozdělení.

2. Úvod

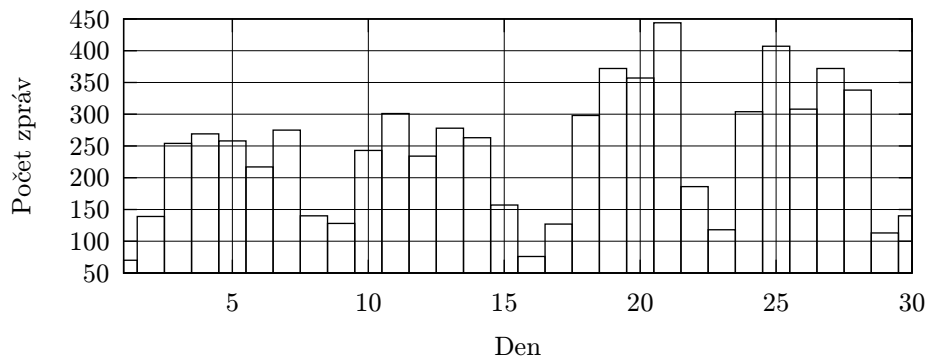
Data k testování rovnoměrného rozdělení byla získána z mailing listu `linux-kernel`, určeného pro vývojáře jádra operačního systému Linuxu.

Počty zpráv přijatých během měsíce dubna 2006 byly vydolovány za pomoci `wget`, `bash`, `grep` a trochy `gcc` z webového archivu na adrese:

<http://www.ussg.iu.edu/hypermail/linux/kernel/>

Celkem bylo přijato 7186 zpráv, počet zpráv v archivu konference pro dané dny:

Den	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.
Zpráv	70	139	254	269	258	217	275	140	128	243	301	234	278	263	157
Den	16.	17.	18.	19.	20.	21.	22.	23.	24.	25.	26.	27.	28.	29.	30.
Zpráv	76	127	298	372	357	444	186	118	304	407	308	372	338	113	140



3. Řešení

Pro testování, zda má hypotéza dané rozdělení použiji χ^2 -test dobré shody [1]:

$$T := \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i},$$

- n – rozsah výběru, počet zpráv přijatých za měsíc $n = 7186$,
- n_i – četnosti jevu, počet zpráv přijatých v daném dni, ($i = 1, \dots, k$),
- k – počet hodnot, tj. počet dní v měsíci dubnu $k = 30$
- p_i – pravděpodobnosti jevu, že nastal právě v daném dni $p_i = 1/30$,
- np_i – teoretické četnost, $np_i = 7186/30 = 239,53$.

Pro $n \rightarrow \infty$ se blíží k $\chi^2(k-1)$. Dosažená významnost: $1 - F_{\chi^2(k-1)}(t)$.
Hypotézu H_0 zamítáme pro $t > Q_{\chi^2(k-1)}(1 - \alpha)$ tj. $1 - F_{\chi^2(k-1)}(t) < \alpha$.

i	n_i	$\frac{(n_i - np_i)^2}{np_i}$	i	n_i	$\frac{(n_i - np_i)^2}{np_i}$
1	70	119,99	16	76	111,65
2	139	42,19	17	127	52,87
3	254	0,87	18	298	14,27
4	269	3,62	19	372	73,26
5	258	1,42	20	357	57,61
6	217	2,12	21	444	174,53
7	275	5,25	22	186	11,96
8	140	41,36	23	118	61,66
9	128	51,93	24	304	17,35
10	243	0,05	25	407	117,08
11	301	15,77	26	308	19,57
12	234	0,13	27	372	73,26
13	278	6,18	28	338	40,48
14	263	2,30	29	113	66,84
15	157	28,44	30	140	41,36

$$T := \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} = \sum_{i=1}^{30} \frac{(n_i - 239,53)^2}{239,53} = 1255,38$$

Hypotézu H_0 zamítneme na hladině významnosti 95% za předpokladu, že platí $T > \chi_{0,95}^2(k-1)$, tedy $T > \chi_{0,95}^2(29) = 42,557$ [2].

Výsledek: $1255,38 > 42,56 \Rightarrow$ Zamítám hypotézu H_0 .

4. Závěr

Hypotéza H_0 předpokládající, že rozložení počtu denních zpráv přijatých do emailové konference během měsíce dubna má rovnoměrné rozdělení byla *zamítnuta*. Může za to především výrazné opadnutí aktivity pisatelů v období víkendů.

Jelikož mi to nedalo, otestoval jsem podobným způsobem ještě hypotézu H'_0 : rozložení počtu denních zpráv *ve všední dny* během měsíce dubna má rovnoměrné rozdělení. Výsledek $T = 333,65$, $T > \chi_{0,95}^2(19) = 30,144$. Na základě tohoto výsledku zamítám i hypotézu H'_0

Reference

- [1] Mirko Navara: *Matematika pro výpočetní techniku*
http://cmp.felk.cvut.cz/~navara/MVT/MVT_pr.pdf

- [2] *Distribution Tables*, <http://www.statsoft.com/textbook/sttable.html>