

Matematika 6F

Semestrální práce – testování hypotéz

1. Zadání:

K hostingu mých webových stránek jsou poskytovány podrobné statistiky návštěvností. Otestuji, zda trvání návštěv www stránek má exponenciální rozložení.

Tedy: **H₀**: Trvání návštěv má exponenciální rozložení.

H₁: Trvání návštěv nemá exponenciální rozložení.

Naměřené hodnoty:

Trvání návštěv	Počet návštěv	Procenta
0s-30s	674	68.5 %
30s-2mn	140	14.2 %
2mn-5mn	73	7.4 %
5mn-15mn	53	5.3 %
15mn-30mn	23	2.3 %
30mn-1h	17	1.7 %
1h+	3	0.3 %

2. Řešení

K testování použijeme χ^2 test dobré shody.

Nejprve je potřeba určit parametr(y) rozložení změřené náhodné veličiny. Jelikož předpokládáme exponenciální rozložení, je nutné učit pouze parametr λ . Tento parametr odhadneme pomocí metody momentů:

$$\lambda = \frac{1}{EX} = \frac{1}{164} = 0,0061$$

Naměřené hodnoty jsou již rozděleny do intervalů potřebných pro χ^2 test dobré shody. Musíme však udělat úpravu. Četnost ve třídě 1h+ je menší než 5, proto tuto třídu sloučíme s 30mn-1h a nezveme ji 30mn+.

Výpočet hodnoty statistiky:

Trvání návštěv	p_i	$np_i = e_i$	o_i	$(o_i - e_i)^2 / e_i$
0s-30s	0,17	164,48	674	1578,35
30s-2mn	0,35	345,96	140	122,62
2mn-5mn	0,32	315,05	73	185,96
5mn-15mn	0,16	153,46	53	65,77
15mn-30mn	0,00	4,03	23	89,38
30mn+	0,00	0,02	20	24003,77
Celkem				26045,85

Hypotézu zamítneme na hladině významnosti α , pokud je hodnota testové statistiky větší než $(1-\alpha)$ -kvantil rozdělení χ^2 o 4 stupních volnosti.

Počet stupňů volnosti jsem určil ze vztahu $k-1-q$, kde k je počet tříd do kterých jsou rozdělena vstupních data (tj. 6), q je počet odhadnutých parametrů, v tomto případě byl odhadnut pouze parametr λ .

Zvolíme-li hladinu významnosti $\alpha = 5\%$, dostáváme $\chi_{0,95}^2(4) = 9,4877$.

Hypotézu H_0 zamítneme na hladině významnosti α jestli-že

$$\sum_{i=1}^k \frac{(o_i - np_i)^2}{np_i} > \chi_{1-\alpha}^2(k-1)$$

Tato nerovnost je splněna, proto zamítáme nulovou hypotézu a přijímáme alternativní, tj. **Trvání návštěv nemá exponenciální rozložení.**

3. Závěr

Zamítnutí nulové hypotézy je pro mne překvapením. Pokud z naměřených hodnot vyneseme závislost do grafu má exponenciální průběh. Proto jsem zvolil exponenciální rozdělení.

4. Použitá literatura

[1] Vladimír Rogalewitz: *Pravděpodobnost a statistika pro inženýry*, První vydání.