



MATEMATICKO-FYZIKÁLNÍ
FAKULTA
Univerzita Karlova

Jaromír Macoun

Testy náhodných bloků

19. října 2021

Oborový seminář

1 Friedmanův test

- Motivační příklad
- Model testu

2 Andersonův-Kannemannův test

- Motivační příklad
- Model testu

3 Porovnání testů

- G_0 distribuční funkce testové statistiky za nulové hypotézy
- Q_x realizace testové statistiky Q
- testy na hladině 5%

Motivační příklad

- Máme 15 lidí, kterým byla podána citrónová šláva, a následně podrobeni Danielovu Inteligenčnímu testu.
- Měříme tep před podáním, během testu a po testu.

	Počty tepů za minutu		
	t=1	t=2	t=3
1.	70	64	69
2.	66	63	61
3.	95	82	80
:	:	:	:
14.	83	72	86
15.	104	99	86

Tabulka: Naměřené hodnoty

- Jak testovat nezávislost na tom, v jaký moment měříme tep?

- Jak testovat nezávislost na tom, v jaký moment měříme tep?
- Lze použít ANOVA?

- Jak testovat nezávislost na tom, v jaký moment měříme tep?
- Lze použít ANOVA?
- Friedmanův test, Andersonův-Kannemanův test

- Jak testovat nezávislost na tom, v jaký moment měříme tep?
- Lze použít ANOVA?
- Friedmanův test, Andersonův-Kannemanův test
- neparametrické testy

- $i \in \{1, \dots, I\}$ počet náhodných bloků
- $j \in \{1, \dots, J\}$ počet léčeb, časový okamžik měření
- Y_{ij} nezávislé náhodné veličiny se spojitou distribuční funkcí F_{ij}
- pro i fixní; R_{ij} pořadí mezi (Y_{i1}, \dots, Y_{iJ})

- $i \in \{1, \dots, I\}$ počet náhodných bloků
- $j \in \{1, \dots, J\}$ počet léčeb, časový okamžik měření
- Y_{ij} nezávislé náhodné veličiny se spojitou distribuční funkcí F_{ij}
- pro i fixní; R_{ij} pořadí mezi (Y_{i1}, \dots, Y_{iJ})
- $I = 15$ počet lidí v experimentu
- $J = 3$ časové okamžiky

- $H_0 : F_{ij}$ nezávisí na j
- Za H_0 (R_{i1}, \dots, R_{iJ}) je rovno libovolné permutaci se stejnou pravděpodobností.
- Testová statistika

$$O = \frac{12}{IJ(J+1)} \sum_{j=1}^J \left(\sum_{i=1}^I R_{ij} - \frac{1}{2} I(J+1) \right)^2$$

- za H_0 při $I \rightarrow \infty$ asymptotické rozdělení $O \sim \chi_{J-1}^2$
- P-hodnota $p = 1 - G_0(O_x)$

Friedmanův test

	Počty tepů za minutu		
	t=1	t=2	t=3
1.	3	2	1
2.	3	2	1
3.	3	2	1
:	:	:	:
14.	2	1	3
15.	3	2	1

Tabulka: Seřazené hodnoty

Motivační příklad

- porovnání 3 druhů fazolí v obchodě
- (levné/normální/drahé)
- Je mezi fazolemi rozdíl?
- Jsou drahé fazole nejlepší?
- 123 lidí ohodnotí známkami

Typ fazolí	Pořadí		
	1.	2.	3.
levné	42	64	17
normální	31	16	76
drahé	50	43	30

Tabulka: Incidence

- $J = 3$; seřaditelných předmětů
- $I = 123$; testovacích objektů (lidí)
- Lze použít χ^2 -test nezávislosti v kontingenční tabulce?

- $J = 3$; seřaditelných předmětů
- $I = 123$; testovacích objektů (lidí)
- Lze použít χ^2 -test nezávislosti v kontingenční tabulce?
- člověk si náhodně vybere posloupnost o délce J
- levné/drahé/normální \implies jedna z $3!$ možností

- n_{ij} -počet j -tých umístění i -tého prvku
- H_0 :
každý prvek má stejnou pravděpodobnost získání daného pořadí
- testová statistika:

$$\frac{J-1}{J} Q^2 = \frac{J-1}{J} \sum_{j=1}^J \sum_{i=1}^J \frac{(n_{ij} - I/J)^2}{I/J}$$

- kritický obor: $[\chi^2_{(J-1)^2}(1-\alpha), \infty)$

- $Y_h =:$ počet lidí, kteří si vyberou h -tou posloupnost
- $\sum_{h=1}^{J!} Y_h = I$
- $Y := (Y_1, \dots, Y_{J!})^T$
- $Y \sim Mult(I, p_1, \dots, p_{J!})$
- za H_0 každá volba posloupnosti pravděpodobnost $p = 1/J!$
- $\mathbb{E} Y_h = Ip; var(Y_h) = Ip(1 - p); cov(Y_h, Y_k) = -Ip^2$

- každá n_{ij} lze vyjádřit jako součet Y_h
- $\mathbb{E} n_{ij} = I/J; \text{var}(n_{ij}) = \frac{I(J-1)}{J^2}$
- Za H_0

$$X_{ij} := \frac{n_{ij} - I/J}{\sqrt{I(J-1)/J^2}} \xrightarrow{d} N(0, 1)$$

- testová statistika lze vyjádřit jako součet kvadrátů X_{ij}

$$Q^2 = \sum_{i=1}^J \sum_{j=1}^J \frac{J-1}{J} X_{ij}^2$$

- *Problém:* za H_0 jsou X_{ij} korelované
- $r := cov(X_{ij}, X_{is}) = \frac{-1}{J-1}$ pro $s \neq j$
- $cov(X_{ij}, X_{ts}) = \frac{1}{(J-1)^2}$ ($= r^2$ pro $s \neq j, i \neq t$)
- lze sestavit kovarianční matici R

- Na data, u kterých používáme Friedmanův test, lze použít po drobné modifikaci i A-K test.
- A-K je obecnější test
- Friedmanův test není konzistentní

Převod Friedmanova testu

	Počty tepů za minutu		
	t=1	t=2	t=3
1.	3	2	1
2.	3	2	1
3.	3	2	1
:	:	:	:
14.	2	1	3
15.	3	2	1

Tabulka: Seřazené hodnoty

	1.	2.	3.
t=1	1	5	9
t=2	9	5	1
t=3	5	5	5

1. Motivační příklad

- *Friedmanův test*
- hodnota testové statistiky $O_x = 8.533$
- kritický obor $(5.99, \infty)$
- P-hodnota 0.014

1. Motivační příklad

- *Friedmanův test*
- hodnota testové statistiky $O_x = 8.533$
- kritický obor $(5.99, \infty)$
- P-hodnota 0.014
- *A-K test*
- hodnota testové statistiky $\frac{J-1}{J} Q_x^2 = 8.533$
- kritický obor $(9.49, \infty)$
- P-hodnota 0.0739

2. Motivační příklad

Typ fazolí	Pořadí		
	1.	2.	3.
levné	42	64	17
normální	31	16	76
drahé	50	43	30

Tabulka: Incidence

2. Motivační příklad

- *A-K test*
- hodnota testové statistiky $\frac{J-1}{J} Q_x^2 = 53.04$
- kritický obor $(9.49, \infty)$
- P-hodnota < 0.0001

Děkuji za pozornost