CVIČENÍ 6

27. březen 2017

Cvičení 1. Ze stránky `http://www.statsci.org/data/general/fullmoon.txt` získejte data `fullmoon`, zopakujte si, co znamenají jednotlivé proměnné a jaké jsou mezi nimi vztahy. Nafitujte v `R` model $\mathcal{F}_{H_1}$ pro závislost počtu pacientů pohotovostní služby psychiatrické kliniky na měsíci v roku (nazvěme jej `model.year`). Zopakujte si interpretaci parametrů $\beta_0, \beta_1, \ldots, \beta_{11}$.

Cvičení 2. Podívejte se na fitované střední hodnoty denního počtu pacientů pohotovostní služby psychiatrické kliniky v jednotlivých měsících seřazené chronologicky.

(a) Uvažujte, jestli by mělo smysl modelovat závislost středních hodnot na měsících uvažovaných jako spojitou proměnnou.

(b) Rozmyslete si, jakou funkční závislost byste v takovém případě zvolili a nafitujte vybraný model.

(c) Jaká funkční závislost by vedla na model ekvivalentní modelu `model.year`?

(d) Bylo by možné některé hypotézy o funkční závislosti testovat i přímo v rámci modelu `model.year`?

Cvičení 3. Zaměříme se na párová porovnávání (rozdíly mezi středními hodnotami pro jednotlivé dvojice měsíců). Zkonstruujte konfidenční intervaly a testy pro rozdíly středních hodnot pomocí

(a) Fisherovy LSD metody,

(b) Scheffého metody,

(c) Tukeyho HSD metody.

Zopakujte si teoretické zdůvodnění jednotlivých metod a z toho odvoďte očekávání o jejich chování v praxi. Graficky porovnejte konfidenční intervaly získané na základě jednotlivých metod a zamyslete se nad tím, jestli je jejich chování na datech `fullmoon` v souladu s vašimi očekáváními.

Tip: Můžete si pomoct přednáškou z týdne 7 z podzimního semestru a přednáškami z jarního semestru.

Cvičení 4.

(a) Uvažujte náhodné jevy $A_1, \ldots, A_k$. Omezte pravděpodobnost sjednocení $\bigcup_{i=1}^{k} A_i$ shora a odvoďte platnost Bonferroniho korekce p-hodnoty při mnohonásobném porovnávání.

(b) Uvažujte nezávislé náhodné jevy $A_1, \ldots, A_k$. Spočtěte pravděpodobnost průniku $\bigcap_{i=1}^{k} A_i$ a odvoďte platnost Šidákovy korekce p-hodnoty při mnohonásobném porovnávání.

Cvičení 5. Spočtěte p-hodnoty a konfidenční intervaly ze Cvičení 3 pomocí Bonferroniho metody a srovnejte výsledky s těmi, které jste dostali ve Cvičení 3.

Domácí úloha (*10 bodů*)

Při testování m hypotéz se setkáváme se situací, kterou můžeme zapsat pomocí následující tabulky.

	Nesignifikantní	Signifikantní	
Pravdivá H_0	U	V	m_0
Nepravdivá H_0	T	S	$m - m_0$
	$m - R$	R	m

Číslo m známe a $m - R$ neznáme, náhodnou veličinu R pozorujeme, zatím co náhodné veličiny U, V a S nepozorujeme. Dvě nejčastěji kontrolovaná kritéria při mnohonásobném testování můžeme pomocí značení z tabulky zapsat jako $FWER = \mathsf{P}(V \geq 1)$ a $FDR = \mathsf{E}(Q)$, kde $Q = V/R$, je-li $R > 0$ a $Q = 0$ jinak. Různé metody mnohonásobného porovnávání jsou většinou navrženy tak, aby kontrolovaly $FWER$ nebo FDR.

(a) Dokažte, že je-li $m_0 = m$, pak $FDR = FWER$. Z toho plyne, že metody mnohonásobného porovnávání kontrolující FDR kontrolují také $FWER$ v slabém smyslu (t.j. za předpokladu, že všechny nulové hypotézy jsou pravdivé).

(b) Dokažte, že je-li $m_0 < m$, pak $FDR \leq FWER$. Z toho plyne, že metody mnohonásobného porovnávání kontrolující $FWER$ kontrolují také FDR. Naopak, metody mnohonásobného porovnávání kontrolující FDR nekontrolují $FWER$ v silném smyslu (t.j. bez ohledu na to, kolik nulových hypotéz je pravdivých). Na druhou stranu mají tyto metody potenciál pro větší sílu než ty, které kontrolují $FWER$ v silném smyslu, jelikož rozdíl mezi veličinami, jejíchž střední hodnoty se v definicích $FWER$ a FDR počítají, lze očekávat tím větší, čím větší je počet nulových hypotéz, které neplatí.

Bonferroniho metoda kontroluje $FWER$ v silném smyslu a Šidákova metoda kontroluje $FWER$ v silném smyslu za předpokladu nezávislosti jednotlivých testových statistik.

K populárním metodám kontrolujícím FDR patří Benjaminiho–Hochbergova metoda. Při ní se získané p-hodnoty seřadí od nejmenší po největší. Je-li $P_{(i)}$ i-tá nejmenší p-hodnota a je-li k největší i, pro které platí, že $P_{(i)} \leq i/m \times q$, pak zamítneme k hypotéz s nejmenšími p-hodnotami. Benjaminiho–Hochbergova metoda kontroluje FDR na hladině q v silném smyslu za předpokladu nezávislosti testových statistik.

U Benjaminiho-Hochbergovy metody jde o takzvanou step-down proceduru. Začneme tím, že vezmeme největší p-hodnotu a je-li ta menší než q, zamítneme všechny hypotézy, jako bychom dělali testování bez úpravy na mnohonásobná porovnávání. Je-li větší než q, podíváme se na druhou největší p-hodnotu a v nejhorším případě postupujeme až k nejmenší p-hodnotě. Příslušnou hypotézu zamítneme, je-li tato menší než q/m, jako bychom používali Bonferroniho korekci.

Jinou populární step-down procedurou je Hochbergova metoda. Ta postupuje podobně jako Benjaminiho-Hochbergova, ale místo hranice $i/m \times q$ srovnává i-tou nejmenší p-hodnotu s hranicí $i/(m + 1 - i) \times \alpha$. Hochbergova metoda kontroluje $FWER$ na hladině α v silném smyslu za předpokladu nezávislosti testových statistik. Není těžké si rozmyslet, že zamítne nejvýše tolik hypotéz co Benjaminiho–Hochbergova metoda.