



Prikupljanje i obradba podataka

prof. dr. sc. Mladen Petrovečki

ožujak 2010.

Znanstveno istraživanje

- znanstveno djelovanje
 - privremeno tumačenje
 - hipoteza
 - zaključak
- provjerljivost
 - tvrdnja
 - obrazloženje
 - znanje



Znanstveni postupci

- postavke
 - ⌘
- činjenice
 - ⌘
- znanje

Neznanstveni postupci

- ustrajnost
(navika, stav, vjerovanje, inercija)
- autoritet
- intuicija (očiglednost)

Istraživačka logika

- deterministički model sustava
- probabilistički model sustava
- vjerojatnost događaja $\Rightarrow p(D)$

$$0 \leq p(D) \leq 1$$

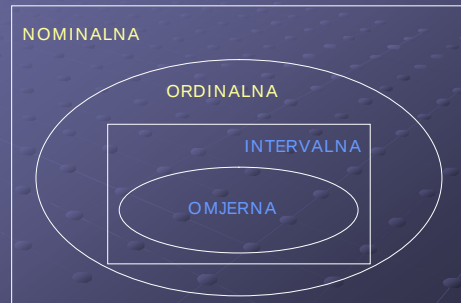
Zašto istraživati?



Odabir pokazatelja

- sve pokazatelje istraživanja
- što više pokazatelja
- pitanje kraja istraživanja
- jednostavni → složeni (podatci)
- mjerne ljestvice ⇒

Ljestvice mjerenja



Statistička hipoteza

- elementarna tvrdnja
- točna (istinita) ili netočna (neistinita)
- provjera hipoteze ⇒ traženje istine



Statistička hipoteza

- istina ⇒ stvarno, objektivno stanje
- probabilistički sustav:
istina ⇒ vjerojatnost
- značajno ⇒ ono što se ostvaruje na svaki drugi način osim slučajno:
iskaz vjerojatnosti ⇒ razina značajnosti

Testiranje statističke hipoteze

hrana u kantini	studenti iz Zagreba	studenti izvan Zagreba
dobra	19 (25%)	21 (28%)
loša	8 (11%)	27 (36%)
ukupno	27	48

Razlikuje li se mišljenje o kakvoći hrane među studentima?
(3 dana x 25 prvih na redu, N=75)

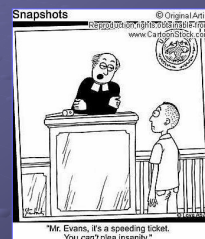
Testiranje statističke hipoteze

- postavljanje hipoteze
- odabir statističkog testa
- određivanje razine značajnosti
- izračunavanje statistike testa
- zaključivanje

1. Postavljanje hipoteze

- nulta – H_0 -hipoteza: razlike nema!
"Nema razlike u mišljenju."
- suprotna – H_1 -hipoteza: razlike ima!
"Razlika u mišljenju postoji."
- samo jedna može biti istinita
- samo jedna može biti prihvaćena, dok će ona druga biti odbačena

Nulta hipoteza



Nema razlike $\approx$ "Nisam kriv"

2. Odabir testa

- testovi dokazivanja statističkih hipoteza
- ovisnost
 - osobine obilježja – mjerne ljestvice
 - osobine uzoraka
 - ♦ veličina
 - ♦ povezanost
 - osobine raspodjele
 - ♦ parametrijski
 - ♦ neparametrijski
 - broj obilježja – uni/bi/multivarijantni testovi

(nastavak)

Ljestvica	Jedan uzorak	Dva uzorka		Tri i više uzoraka	
		Ovisni	Neovisni	Ovisni	Neovisni
Nominalna	binomni test hi-kvadrat	McNemara	Fisher hi-kvadrat/	Cohran	hi-kvad
Ordinalna	Kol/Smir. homologni	Wilcoxon	MW Moses	Friedman	p/medijan KW
Intervalna					
Omjerna	...				

3. Određivanje razine značajnosti

- ♦ P (engl. *probability*)
- ♦ α ako se određuje prije izračunavanja
- ♦ α – vjerojatnost odbacivanja H_0 kad je ona stvarno točna i istinita
- ♦ tzv. α -pogrješka (pogrješka I. vrste)
- ♦ što manja!
- ♦ uobičajene vrijednosti
 - npr. $p < 0,05$

(pogrješke testiranja hipoteze)

PRAVO STANJE		ZAKLJUČENO
RAZLIKA POSTOJI (H_1)	RAZLIKA NE POSTOJI (H_0)	
ISPRAVAN ZAKLJUČAK	α pogreška (I. vrste) (H_0 odbac.)	RAZLIKA POSTOJI
β pogreška (II. vrste)	ISPRAVAN ZAKLJUČAK (H_0 prihv.)	RAZLIKE NEMA

- zašto $P < 0,05$?

POKUS

bacanje novčića: pismo/glava

- $2x$ isto uzastopce = 0,5
- $3x = 0,25$
- $4x = 0,125$
- $5x = 0,063$
- $6x = 0,031$
- $7x = 0,016$
- $8x = 0,008$
- ...

4. Izračun

- matematički račun
- $P \Rightarrow$ egzaktna vrijednost
 - 3 decimalna mjesta
 - npr. $P = 0,325$

5. Zaključivanje

- mala vrijednost $P \Rightarrow$ mala vjerojatnost da odbacujemo istinitost
- zaključivanje:
 - $P < \alpha$
 - vjerojatnost istinitosti H_0 je vrlo mala
 - odbacujemo (ne prihvaćamo) nultu hipotezu
 - prihvaćamo alternativnu hipotezu, H_1
 - potvrdimo je, iskažemo je, uz $P = \dots$

(nastavak – zaključak)

hrana u kantini	studenti iz Zagreba	studenti izvan Zagreba
dobra	19 (25%)	21 (28%)
loša	8 (11%)	27 (36%)
ukupno	27	48

$$\chi^2=3,91, df=1, P=0,048$$

Uzorak i populacija



Razlikuje li se mišljenje o kakvoći hrane među studentima?
(5 dana x 15 prvih na redu u omjeru 1:2)

hrana u kantini	studenti iz Zagreba	studenti izvan Zagreba
dobra	10 (13%)	31 (42%)
loša	15 (20%)	19 (25%)
ukupno	25	50

$$\chi^2=2,43, df=1, p=0,126$$

Maskiranje

- ♦ jednostruko
- ♦ dvostruko
- ♦ trostruko
- ♦ četverostruko

Pogrješka temeljne prosudbe



Kahneman D, Tversky A. On prediction and judgement. Oregon Res Inst Bull 1972;371:100.

Postupak	Stanje	
	postoji (npr. bolest)	ne postoji (npr. zdrav)
pozitivan nalaz	ISPRAVNO POZITIVNI (TP)	LAŽNO POZITIVNI (FP)
negativan nalaz	LAŽNO NEGATIVNI (FN)	ISPRAVNO NEGATIVNI (TN)

$\text{Osjetljivost testa} = TP / (TP + FN)$
 $\text{Specifičnost testa} = TN / (FP + TN)$
 $\text{Pozitivna prediktivna vrijednost} = TP / (TP + FP)$
 $\text{Negativna prediktivna vrijednost} = TN / (FN + TN)$

15%
 85%

PROMETNA NEZGODA, NOĆ, VOZAČ POBJEGNE

$p(\text{zeleni}) = ?$ $p = 0,15 (15\%)$

Osjetljivost testa = 80%, Specifičnost testa = 80% "NI TAXI"

$p(\text{zeleni}) = ?$ $p = 1 (100\%)$

OSPORA SE IZJAVA, TESTIRANJE NOĆNE VIDLJIVOSTI
RAZLUČIVANJA BOJA:
PREPOZNAVANJE 80%, GRIJEŠI 20% $p = 0,8 (80\%)$

$p(\text{zeleni}) = ?$

Svjedok vidi	Stvarna boja vozila	
zeleno	12	17
plavo	03	68
ukupno	15	85

Osjetljivost testa = 80%, Specifičnost testa = 80%

Poz. PV = $12 / (12 + 17) = 12 / 29 = 0,41$ $p = 0,41 (41\%)$

Mudrost zaključivanja!

ZNANOST

$p = 0,41 (41\%)$

Hvala na pozornosti

ZNANOST

Prof. dr. Mladen Petrovečki
mladenp@kdb.hr

Katedra za medicinsku informatiku
Medicinski fakultet Sveučilišta u Rijeci
<http://mi.medri.hr>
Klinički zavod za laboratorijsku dijagnostiku
Klinička bolnica "Dubrava", Zagreb
www.kdb.hr/lab