

PROPENSITY SCORE MATCHING.

STATISTIČKI PRAKTIKUM 2

Primjer i ideja

Promatramo rezultate zaposlenika na testu, obzirom na to jesu li pohađali edukaciju ili ne.

edukacija	bez edukacije
14	10
13	15
18	12
15	17
19	14
17	20
⋮	⋮

Što je s ostalim karakteristikama? Uparujemo ispitanike sa sličnim karakteristikama kako bi rezultati bili usporedivi.

Pitanje: ima li tretman utjecaj na rezultat (postoji li razlika među grupama) - Wilcoxonov test rangova s predznacima ili Friedmanov test (uzorci nisu nezavisni jer smo ih uparili).

Propensity Score Matching

PSM je metoda procjenjivanja efekata tretmana na temelju modela koji ocjenjuje vjerojatnost primanja tretmana (na temelju promatranih karakteristika).

Neka su $X_1, \dots, X_k$ varijable poticaja koje utječu na primanje tretmana, X varijabla koja mjeri efekt tretmana te Y binarna varijabla koja označava tretman ($Y \in \{0, 1\}$).

Ideja:

1. Pomoću binarnog modela procijeniti vjerojatnost ulaska u tretman.
2. Svakoj tretiranoj jedinki naći (jedan ili više) par iz netretirane skupini s najsličnijom vjerojatnosti ulaska u tretman.
3. Efekt tretmana uspoređivati na temelju dvaju dobivenih grupa.

Uparivanje u PSM-u

Traženje para u PSM-u može se odvijati sa i bez vraćanja, najčešće korištene metode su

- ▶ metoda najbližeg susjeda
- ▶ caliper metoda (najbliži susjed s radijusom)
- ▶ metoda radijusa
- ▶ Mahalanobis metoda

Primjer

U datoteci `lalde.txt` nalaze se podaci za analizu efekta programa usavršavanja u SAD-u 70-tih godina.

(LaLonde, Robert. 1986. "Evaluating the Econometric Evaluations of Training Programs." *American Economic Review* 76:604-620.).

```
> names(lalde)
[1] "age"      "educ"     "black"    "hisp"     "married"
[6] "nodegr"   "re74"     "re75"     "re78"     "u74"
[11] "u75"      "treat"
```

Zanima nas utjecaj usavršavanja na zaradu ispitanika mjerenu 1978. godine.

Prvo procijenimo pripadni probit model:

```
> model=glm(treat~age+educ+black+hisp+married+  
  nodegr+re74+re75,family=binomial(link=probit),data=lalonde)  
> summary(model)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	7.300e-01	6.536e-01	1.117	0.26406
age	2.876e-03	8.868e-03	0.324	0.74571
educ	-4.403e-02	4.444e-02	-0.991	0.32175
black	-1.403e-01	2.273e-01	-0.617	0.53696
hisp	-5.232e-01	3.087e-01	-1.694	0.09018 .
married	1.029e-01	1.717e-01	0.599	0.54917
nodegr	-5.622e-01	1.943e-01	-2.893	0.00381 **
re74	-1.969e-05	1.568e-05	-1.255	0.20931
re75	3.864e-05	2.679e-05	1.442	0.14916

Procijenimo vrijednosti varijable odaziva Y , tj. vjerojatnost ulaska u program usavršavanja $\hat{Y}$:

```
> p=predict(model, type="response", se.fit = TRUE)$fit
```

Promotrimo stvarne i procijenjene vrijednosti,

```
> x=cbind(lalonde$treat,p)
```

te za svaku osobu koja je ušla u program ($Y = 1$) nađemo osobu koja nije bila u programu ($Y = 0$) s najbližom vjerojatnošću ulaska u program ($\hat{Y}$).

```

> par=matrix(0,nrow=185,ncol=2)
> indeks=186:445

> for(i in 1:185)
{
  m=min(abs(p[186:445]-p[i]))
  par[i,]=c(indeks[abs(p[186:445]-p[i])==m][1],m)
}

```

Sada možemo testirati efekt tretmana na X =zarada 1978 godine (nakon ulaska u program osposobljavanja).

```

> x=lalonde$re78

```

Podijelimo x u dvije skupine - skupinu tretiranih i skupinu njihovih parova.

```

> x1=x[1:185]
> x0=x[par[,1]]

```

Prigodnim testom provjerimo postoji li razlika u mjerenoj varijabli X između tretirane i netretirane skupine (efekt tretmana)

```
> wilcox.test(x1,x0,alternative="greater",paired=T)
```

Wilcoxon signed rank test with continuity correction

data: x1 and x0

$V = 9873$, $p\text{-value} = 0.0001871$

alternative hypothesis: true location shift is greater than 0