Uporabna statistika

Gregor Dolinar

Fakulteta za elektrotehniko
Univerza v Ljubljani

14. januar 2013

Neparametrične metode pri analizi variance, Kruskal - Wallisov test

ANOVA, primerjamo matematična upanja v posameznih razredih

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij},$$

razredov je a, torej $i = 1, \ldots, a$, v vsakem razredu je n_i opazovanj, torej $j = 1, \ldots, n_i$.

- Če so slučajne spremenljivke ϵ_{ij} neodvisne in normalno porazdeljene, potem smo uporabili F-test.
- Če vemo samo, da so slučajne spremenljivke ϵ_{ij} neodvisne in enako zvezno porazdeljenene, potem lahko uporabimo neparametrični test, na primer Kruskal - Wallisov test.

Radi bi preverili, ali so matematična upanja $\mu_1, \ldots, \mu_a$ v različnih razredih različna. Preverili bomo hipotezo

$$H_0 : \mu_1 = \ldots = \mu_a.$$

Naj bo

$$N = \sum_{j=1}^{a} n_i$$

število vseh opazovanj. Oštevilčimo vsa opazovanja od najmanjšega do največjega.
Če je hipoteza

$$H_0 : \mu_1 = \ldots = \mu_a$$

pravilna, so rangi opazovanj po velikosti enakomerno porazdeljeni po razredih.

Naj bo R_{ij} rang opazovanja Y_{ij}.
V primeru, da je H_0 pravilna, je

$$E(R_{ij}) = \frac{N(N+1)}{2N} = \frac{N+1}{2},$$

matematično upanje povprečne vrednosti rangov v posameznem razredu pa

$$E(\overline{R}_{i.}) = \frac{1}{n_i}\sum_{j=1}^{n_i} E(R_{ij}) = \frac{N+1}{2}.$$

Kruskal-Wallisov test meri, koliko se dejanska vrednost $\overline{R}_{i.}$ razlikuje od pričakovane vrednosti $\frac{N+1}{2}$.

Testna statistika je

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{a} n_i \left(\overline{R}_{i.} - \frac{N+1}{2} \right)^2 .$$

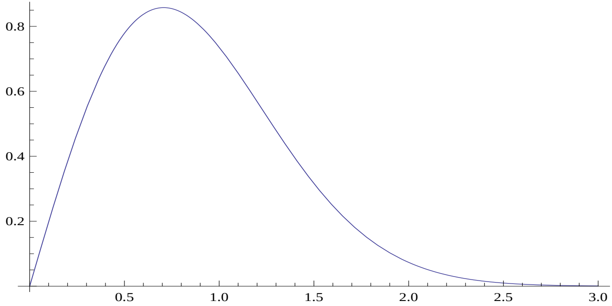
Če je H za nek vzorec velik, potem H_0 zavrnemo.
V primeru, da je $a = 3$ in $n_i \geq 6$ ali $a \geq 3$ in $n_i \geq 5$, potem je H približno χ-kvadrat porazdeljena slučajna spremenljivka z $a - 1$ prostostnimi stopnjami.

Analiza zanesljivosti in življenjske dobe

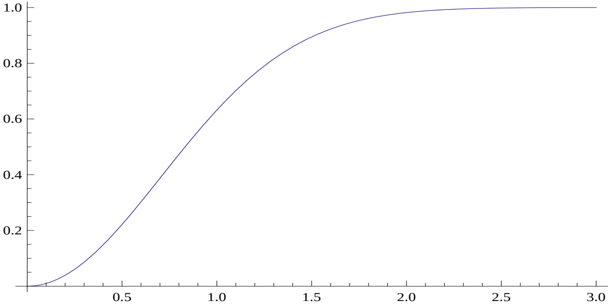
V tehniki je zelo pomembno vprašanje zanesljivost določenega proizvoda. Na primer, zanima nas verjetnost:

- ali bo proizvod deloval v določenem trenutku,
- ali bo proizvod deloval določeno obdobje.

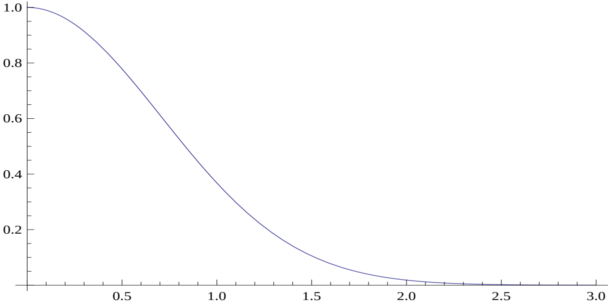
Življenjska doba proizvoda je slučajni dogodek, opišemo jo s funkcijo gostote verjetnosti f, za katero velja $f(t) = 0$ za $t < 0$. Npr., $f(t) = 2te^{-t^2}$ za $t \geq 0$.



Porazdelitvena funkcija $F(t) = \int_0^t f(u)du$ opisuje verjetnost, da je življenjska doba proizvoda med 0 in t, torej je $F(t)$ verjetnost, da proizvod v določenem trenutku t ne bo deloval.



Funkcija zanesljivost R je definirana kot $R(t) = 1 - F(t)$.



Funkcija zanesljivosti $R(t) = 1 - F(t)$ opisuje verjetnost, da je življenjska doba proizvoda daljša od t, torej je $R(t)$ verjetnost, da proizvod v določenem trenutku t bo deloval.

Primer

Denimo, da je funkcija zanesljivosti za nek proizvod $R(t) = \frac{2}{2+t^3}$ v letih.
Koliko naj bo garancijska doba, da bo verjetnost za napako v garancijski dobi manjša od 5 %?
Iščemo tako vrednost t, da bo verjetnost, da bo proizvod v določenem trenutku t deloval, enaka 0.95. Verjetnost, da bo proizvod v določenem trenutku deloval, nam opisuje funkcija zanesljivosti R.
Zapišemo enačbo $R(t) = 0.95$ in dobimo $0.95 = \frac{2}{2+t^3}$.
Rešitev te enačbe je

$$t = \sqrt[3]{\frac{2}{0.95} - 2} = 0.472 \ \ \mathit{leta},$$

kar je enako 5.7 mesecev.

Življenjska doba je slučajna spremenljivka, za katero lahko izračunamo matematično upanje, ki ga v tem primeru imenujemo povprečni čas do okvare (mean time to failure, MTTF).
Torej je

$$MTTF = \mu = \int_0^\infty t f(t) dt.$$

Primer

SSD-diski (solid state drive, tehnologija, kot v USB ključkih, odporni, razvijali so jih predvsem v vojski).
MTTF od 1 do 2 Mhr.
HDD-diski (hard disk drive, uporablja magnetne plošče).
MTTF do 0.5 Mhr.

Ker je $F(t) = \int_0^t f(u)du$, lahko s pomočjo pravila per partes za računanje integralov izpeljemo

$$\begin{aligned}\mu = \int_0^\infty tf(t)dt &= \lim_{M\to\infty}\left(tF(t)|_0^M - \int_0^M F(t)dt\right)\\ &= \lim_{M\to\infty}\left(t(1-R(t))|_0^M - \int_0^M (1-R(t))dt\right)\\ &= \lim_{M\to\infty}\left((t-tR(t)-t)|_0^M + \int_0^M R(t)dt\right) = \int_0^\infty R(t)dt.\end{aligned}$$

Pokazali smo, da je

$$MTTF = \mu = \int_0^\infty tf(t)dt = \int_0^\infty R(t)dt.$$

Torej je MTTF enak ploščini med funkcijo zanesljivosti in osjo x.

Primer

Denimo, da imamo tri proizvode z istim povprečnim časom do okvare, njihove funkcije zanesljivosti pa so različne (glej sliko). Primerjajmo lastnosti teh treh proizvodov.