Eksploracja Danych

Testowanie Hipotez

(c) Marcin Sydow

Wprowadzenie

Testy statystyczne: pocz. XVII wieku (prace J.Arbuthnotta, liczba urodzeń noworodków obu płci w Londynie)

Testowanie hipotez: Karl Pearson (pocz. XX w., testowanie zgodności, test χ^2), potem Egon Pearson (syn) i Jerzy Neyman.

2 przeciwstawne hipotezy:

Hipoteza zerowa

Hipoteza alternatywna

Test: decyzja o przyjęciu lub odrzuceniu hipotezy zerowej

Poziom istotności

α: maksymalne akceptowalne prawdopodobieństwo odrzucenia prawdziwej hipotezy zerowej (tzw. błąd I rodzaju)

Typowe wartości: 0.05, 0.01 (im niższy, tym pewniejszy wynik)

p-wartość: minimalny poziom istotności pdjęcia decyzji o odrzuceniu hipotezy zerowej.

jeśli $p < \alpha$ to mamy podstawy do odrzucenia hipotezy zerowej.

Statystyka testowa.

Testowanie zgodności

Testowanie zgodności z danym rozkładem.

Najpopularniejsze:

- testy normalności
- testy jednostajności
- testy zgodności 2 rozkładów

Testy normalności

$H_0 : F \in \{N(\mu, \sigma) : \mu \in R, \sigma \in R_+\}$

gdzie F to nieznany rozkład, pochodzą z niego obserwacje $x \sim F$

Często procedury statystyczne (np. Gaussowska regresja liniowa lub analiza wariancji (ANOVA)) zakładają zgodność pewnych zmiennych z rozkładem normalnym aby zapewnić prawidłowość procesu testowania.

Który test wybrać?

Jest wiele testów.

Każdy test wykrywa nieco inne rodzaje niezgodności. (np. Shapiro-Wilka (shapiro.test), Andersona-Darlinga, etc.)

Każdy test oparty jest często na dodatkowych założeniach, które należy sprawdzić przed użyciem (np. w pakiecie R, czytając dokumentację odpowiedniego testu)

Moc testu

Prawdopodobieństwo odrzucenia hipotezy zerowej, gdy faktycznie jest fałszywa (uniknięcie błędu II rodzaju) (1-prawdopodobieństwo błędu II rodzaju)

Zależy m.in. od:

- liczby obserwacji (im więcej tym lepiej)
- poziomu istotności

Wizualna ocena normalności

Za pomocą wykresu kwantylowego (qqnorm)

Porównanie kwantyli z rozkładu empirycznego z teroetycznymi kwantylami rozkładu normalnego.

Normalny: punkty układają się na prostej (qqline). Także: qqplot

przykład: x = rnorm(1000); qqnorm(x)

Testy zgodności z rozkładem jednostajnym

Można wykorzystać:

- χ^2
- Kołmogorowa-Smirnowa

Powyższych testów można użyć do testowania zgodności innych rozkładów.

Test X^2

Dzieli dziedzinę zmiennej losowej na przedziały i sprawdza zgodność oczekiwanej liczby obserwacji i faktycznie zaobserwowanej.

Wyższa liczba przedziałów:

- wyższa dokładność (subtelność)
- ale niższa moc testu

Prosta "magiczna" reguła: Przynajmniej 10 obserwacji w każdym przedziale.

(chisq.test, argumentem jest tablica liczników obserwacji w poszczególnych przeziałach)

Przykład

```
segments = seq(0,1,0.2)
x = runif(1000)
a = table(cut(x,segments))
chisq.test(a)
```

Testy zgodności 2 prób

$H_0 : F = G$

F, G - dwa porównywane rozkłady (np. empiryczne)

Hipoteza alternatywna:

- dwustronna $F \neq G$
- jednostronna $F > G$
- jednostronna $F < G$

Używane narzędzia:

- Kołmogorowa-Smirnowa
- Wilcoxona
- qqplot (odchylenie od prostej y=x)

Test Kołmogorowa-Smirnowa

(ks.test)

przykład: ks.test(x,"pnorm")

Można podać inne parametry, np. alternatywa: "two-sided" (domyślna w R), "greater", "less"

Test Wilcoxona

Stosuje się np. do wykrywania różnic w parametrze położenia

przykład: wilcox.test(x,y)

Niewrażliwy na różnice skali, ale może wykryć różnice w skośności.

Równość parametrów położenia

$H_0 : \mu_1 = \mu_2 = ... = \mu_k$

$H_A : \exists_{i \neq j} \mu_j \neq \mu_j$

Użwywane narzędzia:

- test t-Studenta (t.test)
- analiza wariancji (anova)
- test Wilcoxona
- Test Kruskala-Wallisa (uogólnienie testu Wilcoxona na więcej niż 2 próby)

Testy t-Studenta i ANOVA zakładają normalność rozkładu. Pozostałe nie mają takiego założenia (są rangowe), zakładają tylko ciągłość rozkładu (unikanie powtarzających się wartości)

Rodzaje hipotez zerowych

- czy wartość średnia równa się danej liczbie: $H_0 : \mu_x = \mu$
- czy wartości średnie dwóch rozkładów różnią się o stałą: $H_0 : \mu_x - \mu_y = \mu$ (dwa warianty: równe wariancje (R: t.test: var.equal=T) i różne wariancje (wtedy dodawana jest tzw. korekta (np. Welcha). Hipotezę o równości wariancji można sprawdzić np. testem F (var.test)
- czy wartości średnie prób sparowanych różnią się o stałą: $H_0 : \mu_{x-y} = \mu$ (paired = T)

Równość parametrów skali

Testowanie np. równości wariancji (odchyleń standardowych)

$H_0 : \sigma_1 = \sigma_2 = ... = \sigma_k$

dla 2 prób:

- var.test (test F, iloraz wariancji)
- ansari.test
- mood.test

Są też testy dla więcej niż 2 prób:

- bartlett.test
- fligner.test
- levene.test

Uwaga: Wiele testów zakłada np. normalność rozkładu (F, Bartlett)

Przykład


```
a = rnorm(100,1,0)
b = rnorm(100,1,1)
c = rnorm(100,1,2)
ansari.test(a,b)
ansari.test(a,c)
bartlett.test(list(a,b,c))
```

Testowanie proporcji

Czy prawdopodobieństwo sukcesu w rozkładach dwumianowych jest równe:

$H_0 : p_1 = ... = p_k$

R: prop.test

przykład:

prop.test(650,1234,p=0.5) (wynik: p-value: prawdopodobieństwo, że faktycznie było jednak prawdopodobieństwo sukcesu 0.5)

Test korelacji

$H_0 : \rho_{x,y} = 0$

- Pearsona (cor.test)
- Kendalla
- Spearmana

,

Przykład: a = rnorm(50)

b = a + rnorm(50)

cor.test(x,y,method=''spearman'')

Przykładowe pytania/zadania/problemy

- Testowanie hipotez
- Poziom istotności, p-wartość, statystyka testowa, moc testu
- Testy normalności, jednostajności, etc.
- Wizualna ocena normalności
- Testy zgodności
- Testy równości parametrów położenia
- Rodzaje hipotez

Dziękuję za uwagę.