Postavitev hipotez

NUJNO!

Milena Kovač

10. januar 2013

Postavitev in preizkušanje hipotez

Hipoteze zastavimo najprej ob načrtovanju preizkusa

Ob obdelavi jih morda malo popravimo

Preizkus hipotez opravimo v treh korakih

1. Preizkusimo, ali je **model** značilen.
2. Preizkusimo, kateri **vplivi** v modelu so značilni
3. Preizkusimo, kateri **nivoji** pri značilnih vplivih se med seboj razlikujejo

Zapis hipoteze

- Hipoteza je vedno sestavljena iz:

  - ničelne hipoteze in
  - alternativne hipoteze

$$H_0 : P_1 - P_2 = 0 \qquad = \text{med pasmama } P_1 \text{ in } P_2 \text{ ni razlike}$$
$$H_1 : P_1 - P_2 \neq 0 \qquad = \text{med pasmama je razlika}$$

- Ničelno in alternativno hipotezo vedno v paru in v tej obliki!

Ničelna hipoteza - skalarna oblika

Primeri:

$$H_0 : P_1 - P_2 = 0$$

$=$ med pasmama P_1 in P_2 ni razlike

$$H_0 : M_i - M_{i'} = 0$$

$=$ med mesecema M_i in $M_{i'}$ ni razlik

$$H_0 : b = 0$$

$=$ regresijski koeficient b ni različen od nič

$$H_0 : b = 9.7$$

$=$ regresijski koeficient b se ne razlikuje od 9.7

$$H_0 : P_1 - P_2 = 0 \wedge P_1 - P_3 = 0$$

$=$ pasmi P_2 in P_3 se ne razlikujeta od pasme P_1

Linearne kombinacije parametrov za sistematske vplive:

- seštevamo in odštevamo parametre

$$P_1 - P_2$$
$$b$$

- množimo in/ali delimo parametre s konstantami

Hipoteze

- **Ničelna hipoteza:** H_0
  - model ni primeren, vpliv nima učinka, ni razlik med posameznimi nivoji
- **Alterativna hipoteza:** H_1 ali H_2 ali H_a
  - vse ostale možnosti
    - ⋆ model je značilen, vpliv ima učinek, med nivoji so razlike
  - samo del ostalih možnosti (večje ali manjše), drugi del je povsem nepomemben
    - ⋆ nova zdravila morajo biti bolj učinkovita od starih

Ničelna hipoteza - matrična oblika

Primeri:

$$H_0 : \mathbf{K}\beta = \mathbf{0}$$
$$H_0 : \mathbf{K}\beta = \mathbf{m}$$

$\mathbf{K}$ - matrika linearnih kombinacij (lokacijskih) parametrov

za matriko uporabimo tudi črko $\mathbf{H}$

pri eni sami kombinaciji uporabljamo vektor $\mathbf{k}'$ ali $\mathbf{h}'$

β - parametri za sistematske vplive (lokacijski parametri)

$\mathbf{0}$ - vektor pričakovanih vrednosti za linearne kombinacije

$\mathbf{m}$ - vektor pričakovanih vrednosti za linearne kombinacije

(tudi od nič različne vrednosti)

Matrika hipotez

$$\mathbf{K} = \begin{bmatrix} \mathbf{k}_1' \\ \mathbf{k}_2' \\ \vdots \\ \mathbf{k}_p' \end{bmatrix}$$

- vsaka vrstica je lahko samostojna hipoteza
- lahko opravimo tudi skupni preizkus

Alternativna hipoteza - skalarna oblika

$H_1 : P_1 - P_2 \neq 0$	med pasmama P_1 in P_2 je razlika
$H_1 : M_i - M_{i'} \neq 0$	med mesecema M_i in $M_{i'}$ so razlike
$H_1 : b \neq 0$	regresijski koeficient b je različen od 0
$H_1 : b \neq 9.7$	regr. koef. b se razlikuje od 9.7
$H_1 : P_1 - P_2 > 0$	razlika med pasmama P_1 in P_2 je večja od 0
$H_1 : M_i - M_{i'} < 0$	razlika med M_i in $M_{i'}$ je manjša od 0
$H_1 : b > 0$	regresijski koeficient b je večji od 0
$H_1 : b < 9.7$	regr. koef. b je manjši od 9.7

Alternativna hipoteza - matrična oblika

Primeri:

$H_1 : \mathbf{K}\beta \neq \mathbf{0}$	linearne kombinacije so različne od 0
$H_a : \mathbf{K}\beta \neq \mathbf{m}$	linearne kombinacije so različne od konstant v vektorju $\mathbf{m}$
$H_1 : \mathbf{K}\beta > \mathbf{0}$	linearne kombinacije so večje od 0
$H_1 : \mathbf{K}\beta > \mathbf{m}$	linearne kombinacije so večje od konstant v vektorju $\mathbf{m}$
$H_1 : \mathbf{K}\beta < \mathbf{0}$	linearne kombimacije so manjše od 0
$H_1 : \mathbf{K}\beta < \mathbf{m}$	linearne kombinacije so manjše od konstant v vektorju $\mathbf{m}$

Potrditev ali zavrnitev hipoteze

- **ničelna in alternativna hipoteza vedno v paru**

Primer I	Primer II
$H_0: P_1 - P_2 = 0$ $H_1: P_1 - P_2 \neq 0$	$H_0: P_1 - P_2 = 0$ $H_1: P_1 - P_2 > 0$

- lahko so bolj sestavljene, a dajemo prednost enostavnim

- se izključujeta:
  - nikoli ne držita obe hkrati
  - lahko je še tretja možnost, ki pa za nas nikakor ni zanimiva

Potrditev ali zavrnitev hipoteze

- **sprejmemo ničelno hipotezo, alternativno zavrnemo**
  - potem nobena linearna kombinacija (npr. razlika) ni značilna,
  - razlik med nivoji ni,
  - ne izvedemo naslednjega koraka pri preizkušanju hipotez ...
- **zavrnemo ničelno hipotezo, sprejmemo alternativno**
  - iščemo pomembne razlike,
  - med primerjavami je najmanj ena značilna (pomembna)
- **ni tretje možnosti**

Regresijski koeficienti in hipoteze

- če so različni od nič ali od konstante (npr. 9.7)

Primer I	Primer II
$H_0 : b = 0$ $H_1 : b \neq 0$	$H_0 : b = 9.7$ $H_1 : b > 9.7$

- če lahko bolj ugnezdene nadomestimo z manj ugnezdenimi ali glavnimi (želimo poenostaviti model)

Primer I	Če drži ta hipoteza, potem
$H_0 : b_1 = b_2 =$ $= b_2 = \cdots = b_p$	razlik med regresijskimi koeficienti ni lahko jih zamenjamo s skupnim b
$H_1 : b_i \neq 0$	ne moremo poenostaviti

Nikoli ne preizkušamo

- **razlik med nivojema dveh različnih vplivov**

- med pasmo P_1 in mesecem M_2 ni razlik
  $H_0 : P_1 - M_2 = 0$
- regresijski koeficient b se ne razlikuje od vpliva pasme P_1
  $H_0 : b - P_1 = 0$ (različne enote!)
- regresijskih koeficientov pri različnih stopnjah polinoma
  $H_0 : b_I - b_{II} = 0$ (različne enote!)

Kombinirane razlike

- Kombiniranih razlik se izogibamo:

  npr: npr. vsota razlike med pasmama P_1 in P_2 in polovici razlike med mesecema M_1 in M_2

$$H_0 : P_1 - P_2 + 0.5 * (M_1 - M_2) = 0$$
$$H_1 : P_1 - P_2 + 0.5 * (M_1 - M_2) \neq 0$$

- zaradi interpretacije

- hipoteze naj imajo vsebinski pomen, smisel

Postavitev matrike $\mathbf{K}$

- matrična oblika nazoren prikaz hipoteze in uporabljena v statističnih paketih
- vsaka vrstica predstavlja eno linearno kombinacijo
- sestavljene hipoteze pri zapletenih strukturah podatkov in kompleksnem modelu (manjkajoči podatki, interakcije...)
- možnosti:
  - napišemo linearno kombinacijo direktno
  - napišemo najprej kombinacijo za parametra, ki jih primerjamo potem pa sestavimo kombinacijo za preizkus

Preizkus razlik med pasmami - možnost I

Primer: mladice treh pasem (P_i) na treh farmah (F_j)

$$y_{ijk} = \mu + P_i + F_j + e_{ijk}$$

Vektor ocen za lokacijske parametre

$$\widehat{\beta}' = \left[\ \widehat{\mu}\ \ \widehat{P_1}\ \ \widehat{P_2}\ \ \widehat{P_3}\ \ \widehat{F_1}\ \ \widehat{F_2}\ \ \widehat{F_3}\ \right]$$

Preizkus razlik med pasmami - možnost I

Možne razlike med pasmami:

$P_1 - P_2$ razlika med prvo in drugo pasmo

$P_1 - P_3$ razlika med prvo in tretjo pasmo

$P_2 - P_3$ je linearna kombinacija prvih dveh razlik

upoštevamo samo linearno neodvisne kombinacije

Matrični zapis prvih dveh linearnih kombinacij:

$$\begin{array}{ccccccc} \widehat{\mu} & \widehat{P_1} & \widehat{P_2} & \widehat{P_3} & \widehat{F_1} & \widehat{F_2} & \widehat{F_3} \end{array}$$

$$\mathbf{K} = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 \end{bmatrix}$$

Preizkus razlik med pasmami - možnost II

Možne razlike med pasmami:

$P_1 - P_2$ razlika med prvo in drugo pasmo
$P_1 - P_3$ razlika med prvo in tretjo pasmo
$P_2 - P_3$ je linearna kombinacija prvih dveh razlik
upoštevamo samo linearno neodvisne kombinacije
izbor je lahko različen

Matrični zapis prve in tretje linearne kombinacije:

$$\begin{array}{cc} & \begin{bmatrix} \widehat{\mu} & \widehat{P}_1 & \widehat{P}_2 & \widehat{P}_3 & \widehat{F}_1 & \widehat{F}_2 & \widehat{F}_3 \end{bmatrix} \\ \mathbf{K} = & \begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Preberimo naslednje linearne kombinacije

$$
\begin{array}{c}
\begin{bmatrix} \widehat{\mu} & \widehat{P_1} & \widehat{P_2} & \widehat{P_3} & \widehat{F_1} & \widehat{F_2} & \widehat{F_3} \end{bmatrix} \\
\mathbf{H} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & -2 \\ 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}
\end{array}
$$

Obrazložitev:

$\begin{bmatrix} 0 & 0 & 0 & 0 & 1 & -1 & 0 \end{bmatrix}$ razlika med prvo in drugo farmo

$\begin{bmatrix} 0 & 0 & 0 & 0 & 2 & 0 & -2 \end{bmatrix}$ dvakratna razlika med 1. in 3. farmo

$\begin{bmatrix} 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$ povprečje treh farm

Odvisne linearne kombinacije

$$\begin{bmatrix} \widehat{\mu} & \widehat{P_1} & \widehat{P_2} & \widehat{P_3} & \widehat{F_1} & \widehat{F_2} & \widehat{F_3} \end{bmatrix}$$

$$\mathbf{H} = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 2 & -2 & 0 & 0 & 0 \end{bmatrix}$$

Obrazložitev:

$$\begin{array}{r} -2 * \begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 0 & 0 \end{bmatrix} \\ +2 * \begin{bmatrix} 0 & 1 & 0 & -1 & 0 & 0 & 0 \end{bmatrix} \\ \hline \begin{bmatrix} 0 & 0 & 2 & -2 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

tretja vrstica je linearna kombinacija prve in druge;

neprimerna matrika za preizkus hipotez

Ocenljivost linearnih kombinacij

- **Sistemi enačb z nepolnim rangom**
  - nimajo rešitve (nezanimivi!) ali
  - imajo neskončno mnogo rešitev (izbiramo mi ali program sam)
- **Poročamo in preverjamo samo ocenljive funkcije**
  - rešitve ocenljivih fukcij so vedno enake!
- **Poročamo in preverjamo samo najnujnejše**
  - samo linearno neodvisne kombinacije
    (matrika $\boldsymbol{K}$ ali $\boldsymbol{H}$ morata imeti polni rang v vrstici)
  - samo en set linearno neodvisnih kombinacij
  - brez pomena si ne izmišljujemo kompliciranih kombinacij

Sistem enačb

$$\begin{bmatrix} n & n_1 & n_2 & n_3 & \Sigma x_{ij} \\ n_1 & n_1 & & & \Sigma x_{1j} \\ n_2 & & n_2 & & \Sigma x_{2j} \\ n_3 & & & n_3 & \Sigma x_{3j} \\ \Sigma x_{ij} & \Sigma x_{1j} & \Sigma x_{2j} & \Sigma x_{3j} & \Sigma x_{ij}^2 \end{bmatrix} \begin{bmatrix} \widehat{\mu} \\ \widehat{A}_1 \\ \widehat{A}_2 \\ \widehat{A}_3 \\ \widehat{b} \end{bmatrix} = \begin{bmatrix} \Sigma y_{ij} \\ \Sigma y_{1j} \\ \Sigma y_{2j} \\ \Sigma y_{3j} \\ \Sigma x_{ij} y_{ij} \end{bmatrix}$$

$$\mathbf{C}\hat{\beta} = \omega$$

Ocenljivost linearnih kombinacij (nadalj.)

- Linearna kombinacija je ocenljiva, če velja:

za vektor

$$\mathbf{k}'\mathbf{C}^{-}\mathbf{C} = \mathbf{k}'$$

za matriko

$$\mathbf{K}\mathbf{C}^{-}\mathbf{C} = \mathbf{K}$$

statistični paketi preverijo ocenljivost linearnih kombinacij

Ocenljivost linearnih kombinacij (nadalj.)

- Linearna kombinacija ni ocenljiva, če velja:

$$\mathbf{K}\mathbf{C}^{-}\mathbf{C} = \mathbf{H}$$

- toda $\mathbf{H}$ iz zgornje enačbe je ocenljiva, ker velja

$$\mathbf{H}\mathbf{C}^{-}\mathbf{C} = \mathbf{H}$$

pred uporabo presodite, če so hipoteze v $\mathbf{H}$ smiselne

Primer

$$\mathbf{C} \qquad\qquad\qquad \mathbf{C}^-$$

$$\begin{bmatrix} 22 & 6 & 8 & 8 \\ 6 & 6 & & \\ 8 & & 8 & \\ 8 & & & 8 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1/6 & & \\ 0 & & 1/8 & \\ 0 & & & 1/8 \end{bmatrix}$$

- Ali lahko napišete model za zgornji primer?
- Dobili smo eno od neskončno mnogo splošnih inverz:
  ... črtali prvo vrstico in prvi stolpec ...

Primer (nadalj.)

$$\mathbf{C}^{-}\mathbf{C} =$$

$$\begin{array}{cc} & \begin{bmatrix} 22 & 6 & 8 & 8 \\ 6 & 6 & & \\ 8 & & 8 & \\ 8 & & & 8 \end{bmatrix} \\ \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1/6 & & \\ 0 & & 1/8 & \\ 0 & & & 1/8 \end{bmatrix} & \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

- Pazimo na vrstni red matrik!

- Matrika služi kot filter za preverjanje hipotez

Primer

$$\mathbf{KC^{-}C} = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0.5 & 0 & -0.5 \\ 0 & 0 & 2 & -2 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0.5 & 0 & -0.5 \\ 0 & 0 & 2 & -2 \end{bmatrix}$$

- Pogoju ocenljivosti smo zadostili
- Lahko uporabili katerokoli drugo splošno inverzo

Utrdimo!

Preberimo in opišimo naslednji zapis!

$$
\mathbf{K} = \begin{array}{c} \begin{bmatrix} \widehat{\mu} & \widehat{P_1} & \widehat{P_2} & \widehat{P_3} & \widehat{F_1} & \widehat{F_2} & \widehat{F_3} \end{bmatrix} \\ \begin{bmatrix} 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 1 \end{bmatrix} \end{array}
$$

Obrazložitev:

$\begin{bmatrix} 0 & 1 & 0 & 0 & -1 & 0 & 0 \end{bmatrix}$ razlika med 1. pasmo in 1. farmo

$\begin{bmatrix} 0 & 0 & 0 & -1 & 0 & 0 & 1 \end{bmatrix}$ razlika med 3. farmo in 3. pasmo

kombinacije so nesmiselne

Sestavimo linearno kombinacijo!

Povprečni učinek pasme P_i:

- srednja vrednost

- vpliv izbrane pasme

- povprečni vpliv treh pasem, ker so pasme na vseh treh farmah

$$E\left(y_{i.}\right) = 1\mu + 1P_i + 1/3\left(F_1 + F_2 + F_3\right)$$

Sestavimo linearno kombinacijo (nadalj.)!

$$E\left(y_{i.}\right) = 1\mu + 1P_i + 1/3\left(F_1 + F_2 + F_3\right)$$

Linearna kombinacija za povprečni učinek pasme P_1

$$\mathbf{k}_1' = \left[\begin{array}{ccccccc} 1 & 1 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array}\right]$$

Linearna kombinacija za povprečni učinek pasme P_2

$$\mathbf{k}_2' = \left[\begin{array}{ccccccc} 1 & 0 & 1 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array}\right]$$

Linearna kombinacija za razliko med P_1in P_2

$$\begin{array}{rrl} \mathbf{k}_1' & +1* & \left[\begin{array}{ccccccc} 1 & 1 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array}\right] \\ -\mathbf{k}_2' & -1* & \left[\begin{array}{ccccccc} 1 & 0 & 1 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array}\right] \\ \hline \mathbf{k}_1' - \mathbf{k}_2' & & \left[\begin{array}{ccccccc} 0 & 1 & -1 & 0 & 0 & 0 & 0 \end{array}\right] \end{array}$$

Linearne kombinacije za testiranje hipotez

- biti morajo smiselne
- moramo jih znati razložiti
- biti morajo ocenljive
- ocenljivost je odvisna od statističnega modela in strukture podatkov (manjkajoči podatki!)
- po vsakem čiščenju ali preurejanju podatkov moramo preveriti ocenljivost
- istočasno preizkušamo samo hipoteze, ki so med seboj linearno neodvisne