

Exercicis pràctic. Test Permutacions 17-18 (aula d'ordinadors)

Sergi Civit,

1 de març de 2018

1 Exercici 1

1.1 Enunciat de la Situació Experimental

L'origen de la civilització etrusca encara és un misteri per als antropòlegs. En concret, una qüestió que es planteja és si eren originaris de la península italiana o si procedien d'un altre lloc. Es va pensar que una forma de respondre a aquesta pregunta seria comparar els actuals italians amb les restes etrusques, mitjançant un estudi antropomètric. Aquí utilitzarem una part petita d'aquest estudi. En concret, es va mesurar, en mil·límetres, l'amplada màxima del crani de 8 restes de barons homes i la de 10 italians, tots ells escollits a l'atzar. Els resultats obtinguts van ser:

etruscs	141	132	154	142	141	150	134	140		
italians	133	138	136	125	135	130	127	131	116	128

1.2 Qüestions a resoldre

A partir d'aquestes dades,

- Es podria concloure **l'existència de diferències significatives entre italians actuals i etruscos a un nivell de significació 0.05?**

Pots suposar que, si hi ha diferències, són degudes al paràmetre de localització (mitjana, mediana...) i no al paràmetre de dispersió, però no hauries de suposar cap forma concreta per a la distribució de l'amplada màxima del crani.

1. Indica, justificadament, què caldria permutar en la present situació.
2. Indica exactament quantes permutacions possibles hi hauria.
3. Prova de permutacions: calcula el p-valor exacte si utilitzem com a estadístic de test la **diferència de mitjanes mostrals**. Per aquestes dades, indica el p-valor i la conclusió final

4. ¿Tindria sentit utilitzar com a estadístic de test solament la suma dels valors dels etruscos?. Justificar la resposta
5. Realitza una prova de permutacions de Monte Carlo per les dades i hipòtesis anteriors. Indica el p-valor obtingut i la conclusió final.

1.3 Material a entregar com a solució dins de la Tasca 1 del campus

1. Pdf o Word amb la resposta, raonada, a les preguntes anteriors.
2. Script R amb els càlculs realitzats per respondre a les preguntes

1.4 Solució

Carreguem les dades i determinem clarament els grups i les seves grandàries mostrals

```
> # Les dades en dos vectors:
> etruscos = c(141, 132, 154, 142, 141, 150, 134, 140)
> italians = c(133, 138, 136, 125, 135, 130, 127, 131, 116, 128)
> cranis = c(etruscos, italians)
> n1 = length(etruscos)
> n2 = length(italians)
> n = c(n1, n2)
> N = sum(n)
```

1.4.1 Prova de permutacions

Caldria "permutar" només un grup (el menys nombros) i recordant que dins del grup no importa l'ordre les **permutacions** passen a ser combinacions de N en grups de $n1$ on aquest $n1$ són els individus del grup etrusc i tindriem un nombre de permutacions raonables per dur a terme el test exacte.

```
> choose(N, n1)

[1] 43758

> # Funció pel càlcul de la diferència de mitjanes:
> diff.means <- function(indexs, vector.dades) {
+   mean(vector.dades[indexs]) - mean(vector.dades[-indexs])
+ }
> # Diferència de mitjanes per les dades originals:
> dMeans = diff.means(1:n1, cranis)
> # Indexos de les combinacions de N elements agafats en grups de n1:
> combs = combn(N, n1)
> # Càlcul de la diferència de mitjanes per cada combinació:
> dMeansPerm = apply(combs, 2, diff.means, vector.dades = cranis)
```

```
> # p-valor del test bilateral:
> sum(abs(dMeansPerm) >= abs(dMeans)) / ncol(combs)

[1] 0.001348325
```

Amb aquestes dades no seria correcte utilitzar com a estadístic pel test de permutacions la suma dels valors d'un dels grups ja que els grups no són balancejats (8 i 10 dades).

Novament la conclusió final és que rebutgem H_0 .

2 Exercici 2

2.1 Enunciat: Situació experimental

(Extret de la plana web del prof. A. Pitarque, Univ. València.) Comparamos dos muestras aleatorias de 10 hombres y de 10 mujeres de edades comprendidas entre los 18 y los 22 años en un ítem que mide su autoestima (escala ordinal de 0 a 10 puntos). Los datos fueron:

HOMBRES	8	7	6	8	7	5	6	4	9	9
MUJERES	8	6	5	6	5	4	4	4	6	4

a) ¿Podemos afirmar que ambas muestras difieren significativamente en autoestima? b) ¿Podemos afirmar que la autoestima de los hombres es significativamente mayor que la de las mujeres?

Tant la pregunta a) com la b) has de respondre-la si té sentit emprant el test de permutacions. Encas que consieris que sigui cert, justifica la resposta i duu a terme el test

2.2 Solució

Clarament la variable de resposta no és contínua. Ens queda el dubte de si es tracta d'una variable quantitativa però discreta (que inclouria el cas purament "ordinal") o una escala únicament ordinal. En aquest darrer cas les diferències entre valors no tindrien sentit, per exemple no estaria clar si entre un valor de 8 i un de 5 hi ha la mateixa diferència que entre un valor de 6 i un de 3. Això faria que molts dels estadístics habituals no fossin aplicables.

La pregunta a) planteja una hipòtesi alternativa bilateral i la b) unilateral.

```
> # Les dades en dos vectors:
> homes = c(8, 7, 6, 8, 7, 5, 6, 4, 9, 9)
> dones = c(8, 6, 5, 6, 5, 4, 4, 4, 6, 4)
> autoestima = c(homes, dones)
> n1 = length(homes)
> n2 = length(dones)
> n = c(n1, n2)
> N = sum(n)
```

2.2.1 Prova de permutacions

Si podem considerar que la variable de resposta és discreta, no hi ha cap problema en basar la comparació en la diferència de mitjanes, com en casos anteriors.

```
> # Funció pel càlcul de la diferència de mitjanes:
> diff.means <- function(indexs, vector.dades) {
+   mean(vector.dades[indexs]) - mean(vector.dades[-indexs])
+ }
> # Diferència de mitjanes per les dades originals:
> dMeans = diff.means(1:n1, autoestima)
> # Indexos de les combinacions de N elements agafats en grups de n1:
> combs = combn(N, n1)
> # Càlcul de la diferència de mitjanes per cada combinació:
> dMeansPerm = apply(combs, 2, diff.means, vector.dades = autoestima)
> # Pregunta a) p-valor del test bilateral:
> sum(abs(dMeansPerm) >= abs(dMeans)) / ncol(combs)

[1] 0.03236701

> # Pregunta b) p-valor del test unilateral:
> sum(dMeansPerm >= dMeans) / ncol(combs)

[1] 0.01618351
```

Novament la conclusió final és que rebutgem H_0 .

Si l'escala de mesura de la variable és purament ordinal, el procés de permutació és igualment vàlid, però és discutible analitzar-lo en termes de mitjanes. Una possibilitat seria treballar amb els rangs.... pero no pas a partir de l'enfoc de permutacions.