EXAMEN PARCIAL, CURS 2013-2014. ***MÈTODES NO PARAMÈTRICS I DE REMOSTREIG*****. GRAU EN ESTADÍSTICA.** *Totes les preguntes puntuen igual. Respon als mateixos fulls de l'examen.*

ID	Age	VSMT	Ranked_VSMT	var
1	1	98	39.0	
2	1	73	25.0	
3	1	41	3.0	
4	1	51	7.5	
5	1	82	30.0	
6	1	66	15.5	
7	1	97	37.0	
8	1	92	34.5	
9	1	74	26.0	
10	1	71	20.5	
11	1	98	39.0	
12	1	43	5.0	
13	1	92	34.5	
14	1	81	28.0	
15	1	65	14.0	
16	1	72	23.0	
17	1	71	20.5	
18	1	72	23.0	
19	1	72	23.0	
20	1	92	34.5	
21	1	66	15.5	
22	1	82	30.0	
23	1	51	7.5	
24	1	86	32.0	
25	1	67	17.5	
26	1	67	17.5	
27	2	41	3.0	
28	2	54	10.0	
29	2	61	13.0	
30	2	98	39.0	
31	2	41	3.0	
32	2	54	10.0	
33	2	69	19.0	
34	2	82	30.0	
35	2	34	1.0	
36	2	92	34.5	
37	2	47	6.0	
38	2	55	12.0	
39	2	54	10.0	
40	2	79	27.0	

Exercici 1. En un estudi sobre la relació entre envelliment i memòria visual, uns investigadors van mesurar la variable VSMT (Visual Spatial Memory Task) en una mostra formada per 26 dones joves (valor "1" de la variable "Age") i 14 dones grans (valor "2"). A la columna "Ranked_VSMT" s'indiquen els rangs dels 40 valors de VSMT. Com es pot apreciar hi ha alguns empats.

L'objectiu de l'estudi era demostrar una disminució del valor de VSMT amb l'edat i estimar aquest grau de disminució. Es va decidir utilitzar un nivell de significació de 0.05 i un nivell de confiança de 0.95.

Utilitzant si cal els llistats al final d'aquest enunciat[1], respon les següents preguntes :

1) Indica el nom d'una prova d'hipòtesis basada en rangs que sigui apropiada per a intentar demostrar l'existència de l'efecte de l'edat esmentat abans. Explica les suposicions que cal fer per a poder considerar vàlida aquesta prova. Enuncia les hipòtesis nul·la i alternativa associades al problema plantejat.

Una prova adequada seria la de Mann-Whitney-Wilcoxon per a 2 grups independents. Caldria suposar que les dades de cada grup són una mostra i.i.d d'una distribució contínua qualsevol, la mateixa per a tots dos grups, excepte potser les medianes poblacionals. La prova també seria vàlida si només es disposés de la informació dels rangs. El contrast d'hipòtesis plantejat seria una hipòtesi nul·la d'igualtat de medianes, H_0: $\mu_{joves} = \mu_{grans}$ en front a una hipòtesi alternativa H_1: $\mu_{joves} > \mu_{grans}$ que afirmaria que la mediana de VSMT és més gran per les joves.

[1] A tots els exercicis, alguns dels càlculs poden no ser necessaris. Has de triar solament els que serveixen per a la solució.

NOM I COGNOMS: **FIRMA:**

2) Calcula el valor de l'estadístic de test de la prova anterior i la seva conclusió final. Ateses les mides mostrals implicades, segurament hauràs d'utilitzar l'aproximació asimptòtica d'aquest test. **Ignora l'existència d'empats**.

Segurament, en vista que cal resoldre aquesta qüestió "a mà" calcularíem l'estadístic U:

$$U = \min\left\{R_{1.} - \frac{n_1(n_1+1)}{2}, R_{2.} - \frac{n_2(n_2+1)}{2}\right\}$$
$$u = \min\left\{602.5 - \frac{26(26+1)}{2}, 217.5 - \frac{14(14+1)}{2}\right\} = 112.5$$

Però, per poc, veiem que les taules del test (com a mínim les entregades a classe) no inclouen completament les mides mostrals 26 i 14. Per tant, haurem d'utilitzar l'aproximació asimptòtica normal, que basarem en l'estadístic W definit com:

$$W = R_{1.} - \frac{n_1(n_1+1)}{2}$$

i que amb aquestes dades pren el valor

$$w = 602.5 - \frac{26(26+1)}{2} = 251.5.$$

Sota H_0 la mitjana i la variància de W serien els valors indicats als llistats: 182 i 1243.667 (aquest darrer ignorant empats), i tindríem $z = (251.5 - 182)/\sqrt{1243.667} = 1.97$. Aquest valor caldria comparar-lo amb el valor crític normal que, si utilitzèssim la taula per dues cues, correspodria a $z_{2\times 0.05} = z_{0.1} = 1.64$. Atès que la hipòtesi alternativa és unilateral i afirma que la mediana per les joves és superior a la de les grans, com que 1.97 > 1.64, rebutjarem H_0.

3) Respon les mateixes qüestions de la pregunta anterior, però **ara tenint en compte l'existència d'empats** i fent les correccions adequades.

Ara la majoria de passos anteriors són vàlids, l'única cosa que hem de fer és corregir la variància per la presència d'empats. Segons els llistats hi ha $s = 10$ sèries d'empats, de llargades t_i = 3, 3, 2, 3, 2, 4, 2, 3, 2 i 3. Segons la fórmula de correcció de la variància, cal restar a la variància que hem fet servir abans el valor:

$$\frac{n_1 n_2\left(\sum_{i=1}^{g}\left(t_i^3-t_i\right)\right)}{12\left(n_1+n_2\right)\left(n_1+n_2-1\right)}=\frac{26\cdot 14\left(\left(3^3-3\right)+\ldots+\left(3^3-3\right)\right)}{12\cdot 40\cdot 39}=$$

$$\frac{74256}{18720}=3.97$$

Per tant el valor de z corregit seria pràcticament igual a l'anterior (la diferència estaria més enllà de les dues primeres xifres decimals), $z=\left(251.5-182\right)/\sqrt{1243.667-3.97}=1.97$ i evidentment la conclusió seria la mateixa.

4) Obté l'estimació puntual i l'interval de confiança, **associats al test anterior**, per a la diferència de medianes entre el grup "1" i el grup "2". (Pel cas de l'interval de confiança **es demana un interval bilateral**, que en realitat estaria associat a una hipòtesi alternativa bilateral al test.)

L'estimació de la diferència de medianes correspondria a la mediana de les 364 possibles diferències entre joves i grans. Per tant, segons el llistat seria 13.

L'extrem inferior i superior de l'interval de confiança correspondrien als valors que, dins el vector de les 364 diferències ordenades, ocuparien respectivament les posicions λ i ν definides a partir dels següents càlculs:

$$\lambda^*=\frac{n_1 n_2}{2}-z_\alpha\sqrt{\frac{n_1 n_2\left(n_1+n_2+1\right)}{12}}=182-1.96\times 35.27=112.99$$

$$\upsilon^*=1+\frac{n_1 n_2}{2}+z_\alpha\sqrt{\frac{n_1 n_2\left(n_1+n_2+1\right)}{12}}=183+1.96\times 35.27=252.01$$

Arrodonint els valors anteriors obtenim λ = 113 i ν = 252. Veiem que sumen 365 (=364 + 1) de manera que l'interval buscat comença a la posició 113 (on hi ha un 0) del vector de les 364 diferències ordenades i acaba a la posició 252 (on hi ha un 27): [0, 27][2].

---

[2] Com que hi ha empats, en realitat caldria calcular $z_\alpha\sqrt{\frac{n_1 n_2\left(n_1+n_2+1\right)}{12}-\frac{n_1 n_2\left(\sum_{i=1}^{g}\left(t_i^3-t_i\right)\right)}{12\left(n_1+n_2\right)\left(n_1+n_2-1\right)}}$ (el resultat anterior està obtingut així) però la diferència és mínima i acaba donant el mateix.

5) Calcula l'interval de confiança bootstrap-t bilateral per a la diferència entre la mitjana de VSMT al grup "1" i la mitjana al grup "2".

Demana l'interval bootstrap-t sense simetritzar (cues iguals). Per tant, hem de buscar els valors crítics "t" que deixin 0.025 de probabilitat a cada cua. Al llistat, com a resultat de `> quantile(t.boot, probs = c(0.025, 0.05, 0.95, 0.975))`, hi trobem els valors t_1 = -1.9182 i t_2 = 2.2800. Com que la diferència de mitjanes a la mostra original val 12.4231 i el seu error estàndard val 6.1399, l'interval serà :

$$\left[12.4231 - 2.2800 \times 6.1399,\ 12.4231 + 1.9182 \times 6.1399\right] = \left[-1.58,\ 24.20\right].$$

6) Si hi hagués més de 2 grups d'edat (per exemple, "adolescents", "joves", "mitjana edat" i "grans"), indica el nom d'una prova de rangs apropiada per a determinar si hi ha diferències en la mediana de VSMT. Si la prova anterior determinés l'existència de diferències, explica com podríem estudiar quins grups són significativament diferents entre ells, utilitzant també una prova basada en rangs i controlant adequadament l'error de tipus I.

La prova adequada seria la de Kruskal-Wallis.

Si aquesta prova determinés que hi ha alguna diferència (és a dir, rebutgèssim la hipòtesi nul·la que no hi ha cap diferència), podríem realitzar proves de Mann-Whitney-Wilcoxon per fer comparacions 2 a 2 de tots els grups d'edat i determinar quins grups són realment diferents entre ells. Per controlar l'error de tipus I i evitar l'efecte de la realització de múltiples tests, hauríem de corregir els p-valors obtinguts amb un mètode adequat. Una bona opció és el mètode de Holm, que no requereix que els p-valors siguin independents (fixem-nos que molts d'ells hauran estat calculats amb part de les dades en comú).

LLISTATS

```
> joves = c(98, 73, 41, 51, 82, 66, 97, 92, 74, 71, 98, 43, 92, 81,
+   65, 72, 71, 72, 72, 92, 66, 82, 51, 86, 67, 67)
> grans = c(41, 54, 61, 98, 41, 54, 69, 82, 34, 92, 47, 55, 54, 79)
>
> Age = factor( c( rep(1, length(joves)), rep(2, length(grans))))
> VSMT = c(joves, grans)
> Ranked_VSMT = rank(VSMT)
> Ranked_VSMT
 [1] 39.0 25.0  3.0  7.5 30.0 15.5 37.0 34.5 26.0 20.5 39.0  5.0 34.5 28.0 14.0
[16] 23.0 20.5 23.0 23.0 34.5 15.5 30.0  7.5 32.0 17.5 17.5  3.0 10.0 13.0 39.0
[31]  3.0 10.0 19.0 30.0  1.0 34.5  6.0 12.0 10.0 27.0
>
> N = length(VSMT)
> N
[1] 40
> n1 = length(joves)
> n1
[1] 26
> n2 = length(grans)
> n2
[1] 14
> n1 * n2 / 2
[1] 182
> n1 * n2 * (n1 + n2 + 1) / 12
[1] 1243.667
>
>
> # Suma i mitjana de rangs dins cada grup d'edat:
> tapply(Ranked_VSMT, Age, sum)
    1     2
602.5 217.5
> tapply(Ranked_VSMT, Age, mean)
       1        2
23.17308 15.53571
>
> # Funció que determina totes les sèries d'empats i la seva llargada per
> # un vector qualsevol 'x':
> ties = function(x) {
+   ti = sapply(lapply(unique(x), function(xi, x) x %in% xi, x),sum)
+   return(ti[ti > 1])
+ }
>
> # Empats a VSMT:
> ti = ties(VSMT)
> ti
 [1] 3 3 2 3 2 4 2 3 2 3
> # Hi ha una primera sèrie de 3 valors empatats, una segona sèrie amb 3
> # valors empatats, una tercera amb 2, etc.
>
> # Calculem totes les possibles diferències entre els valors de VSMT
> # de 'joves' i 'grans' (26x14 = 364 diferències possibles):
> dij = outer(joves, grans, "-")
> # i les ordenem de més petit a més gran:
> sort(dij)
  [1] -57 -55 -51 -49 -47 -47 -41 -41 -41 -39 -38 -36 -33 -32 -32 -31 -31 -31
 [19] -31 -28 -28 -28 -27 -27 -27 -26 -26 -26 -26 -26 -26 -25 -25 -25 -24 -21
 [37] -21 -20 -20 -20 -20 -19 -18 -18 -18 -18 -17 -17 -16 -16 -16 -16 -15 -15
 [55] -14 -14 -13 -13 -13 -13 -13 -12 -12 -12 -12 -11 -11 -11 -11 -11 -11 -10
 [73] -10 -10 -10 -10 -10 -10  -9  -8  -8  -8  -7  -7  -7  -6  -6  -6  -6  -6
 [91]  -6  -5  -4  -4  -4  -4  -3  -3  -3  -3  -3  -3  -3  -3  -2  -2  -1  -1
[109]   0   0   0   0   0   0   0   0   0   2   2   2   2   2   3   3   3   3
[127]   3   4   4   4   4   4   5   5   5   5   6   6   6   6   7   7   9  10
[145]  10  10  10  10  10  10  10  10  10  11  11  11  11  11  11  11  11  12
[163]  12  12  12  12  12  12  12  12  12  13  13  13  13  13  13  13  13  13
[181]  13  13  13  15  16  16  16  16  17  17  17  17  17  17  17  17  17  17
[199]  17  17  18  18  18  18  18  18  18  18  18  18  18  18  19  19  19  19
[217]  19  19  19  19  20  20  20  20  20  20  21  21  23  23  23  24  24  24
[235]  24  25  25  25  25  25  25  25  25  26  26  26  26  26  26  27  27  27
[253]  27  27  27  28  28  28  28  28  28  28  29  29  30  30  30  30  31  31
[271]  31  31  31  31  31  31  31  31  31  32  32  32  32  32  32  32  33  33
[289]  33  33  34  35  35  36  37  37  37  37  37  37  37  38  38  38  38  38
[307]  38  38  38  38  38  38  38  39  39  40  40  40  41  41  41  41  42  43
[325]  43  43  43  43  44  44  44  44  44  44  45  45  45  45  45  47  48  48
[343]  50  51  51  51  51  51  51  51  51  52  56  56  57  57  57  57  58  58
[361]  58  63  64  64
> # Mediana de les 364 diferències:
> median(dij)
[1] 13
>
> # Càlculs Bootstrap
> # Bootstrap sobre les 26+14 dades, com dos grups independents
> # ============================================================
> # Funció que calcula la diferència de les mitjanes de dues mostres x, y:
> difM = function(x, y) mean(x) - mean(y)
> # Funció que calcula l'error estàndard de la diferència de les mitjanes
> # de dues mostres independents x i y:
```

```
> se.difM = function(x, y) sqrt(var(x) / length(x) + var(y) / length(y))
> #
> # Diferència de mitjanes i el seu error estàndard sobre la mostra original:
> dm = difM(joves, grans)
> dm
[1] 12.42308
> se.dm = se.difM(joves, grans)
> se.dm
[1] 6.139879
> #
> #
> nboot = 10000
> # 'nboot' remostres bootstrap no paramètric dels 40 valors,
> # estratificat per separat dins 'joves' i 'grans'.
> # Per cada remostra calculem el valor de la diferència de mitjanes
> # i el seu error estàndard:
> set.seed(321)
> stats.boot = replicate(nboot,
+ {
+   joves.boot = sample(joves, replace = TRUE)
+   grans.boot = sample(grans, replace = TRUE)
+   c(difM(joves.boot, grans.boot), se.difM(joves.boot, grans.boot))
+ }
+ )
>
> rownames(stats.boot) = c("dm", "se.dm")
>
> # stats.boot és una matriu de 2 files i 10000 columnes. La primera fila conté el valor
> # de diferència de mitjanes per a cada remostra bootstrap. La segona fila conté el
> # seu error estàndard.
> # Valors per les primeres 10 remostres bootstrap:
> stats.boot[,1:10]
          [,1]      [,2]     [,3]      [,4]     [,5]     [,6]      [,7]      [,8]
dm    5.785714 11.423077 5.494505 0.4285714 8.417582 9.351648 13.752747 13.736264
se.dm 6.181236  4.827677 7.467477 5.7361724 6.413677 5.885961  5.521271  5.696065
           [,9]     [,10]
dm    21.291209 21.236264
se.dm  6.027298  5.498451
>
> # Estadístic t estudentitzat per cada remostra bootstrap (vector de 10000 valors):
> t.boot = (stats.boot["dm",] - dm) / stats.boot["se.dm",]
> #
> # Alguns quantils de les dades anteriors:
> quantile(stats.boot["dm",], probs = c(0.025, 0.05, 0.95, 0.975))
      2.5%         5%        95%      97.5%
 0.5489011  2.4557692 21.9450549 23.5989011
> quantile(stats.boot["se.dm",], probs = c(0.025, 0.05, 0.95, 0.975))
    2.5%       5%      95%    97.5%
4.410262 4.671071 6.986974 7.172303
> quantile(t.boot, probs = c(0.025, 0.05, 0.95, 0.975))
     2.5%        5%       95%     97.5%
-1.918217 -1.610007  1.842555  2.280021
> quantile(abs(t.boot), probs = c(0.025, 0.05, 0.95, 0.975))
      2.5%         5%        95%      97.5%
0.03206018 0.06084763 2.10543949 2.41572811
```

Exercici 2. Uns estudiants d'una escola de negocis van realitzar un petit estudi sobre les diferències de preu entre els llibres del seu tema adquirits en una llibreria convencional o a través d'Internet. En una gran llibreria van escollir a l'atzar una mostra de 19 llibres dins la secció d'economia i finances, i van prendre nota del seu preu. Per aquests mateixos llibres van buscar ofertes en línia (del llibre nou, no de segona mà), i es van quedar amb la primera que van trobar. La taula de l'esquerra mostra el preu en dòlars de cada llibre, en llibreria ("bookstore") i "online". Per estar molt segurs de les seves conclusions, van decidir utilitzar un **nivell de significació de 0.01** i un **nivell de confiança 0.99**. Utilitzarem aquests valors **en tot l'exercici**.

Author	Title	Bookstore	Online
Pride	*Business* 10/e	132.75	136.91
Carroll	*Business and Society*	201.50	178.58
Quinn	*Ethics for the Information Age*	80.00	65.00
Bade	*Foundations of Microeconomics* 5/e	153.50	120.43
Case	*Principles of Macroeconomics* 9/e	153.50	217.99
Brigham	*Financial Management* 13/e	216.00	197.10
Griffin	*Organizational Behavior* 9/e	199.75	168.71
George	*Understanding and Managing Organizational Behavior* 5/e	147.00	178.63
Grewal	*Marketing* 2/e	132.00	95.89
Barlow	*Abnormal Psychology*	182.25	145.49
Foner	*Give Me Liberty: Seagull Ed.* (V2) 2/e	45.50	37.60
Federer	*Mathematical Interest Theory* 2/e	89.95	91.69
Hoyle	*Advanced Accounting* 9/e	123.02	148.41
Haviland	*Talking About People* 4/e	57.50	53.93
Fuller	*Information Systems Project Management*	88.25	83.69
Pindyck	*Macroeconomics* 7/e	189.25	133.32
Mankiw	*Macroeconomics* 7/e	179.25	151.48
Shapiro	*Multinational Financial Management* 9/e	210.25	147.30
Losco	*American Government 2010 Edition*	66.75	55.16

Suposem que l'objectiu del seu estudi era **detectar diferències** de preu, sense (com a mínim inicialment) cap idea preconcebuda sobre el signe d'aquestes diferències.

Utilitzant quan calgui els llistats adjunts, respon les següents preguntes[3]:

1) Per una prova d'hipòtesis basada en rangs apropiada per a intentar demostrar l'existència de les diferències de medianes indicades abans, calcula l'estadístic de test i indica la conclusió final.

Atès que es tracta de dades aparellades, una prova de rangs adequada seria la de Wilcoxon per rangs amb signe. Es planteja una prova bilateral: hipòtesi nul·la d'igualtat de medianes i alternativa de diferència.

Només tenen signe negatiu les diferències primera, cinquena, vuitena, onzena i dotzena. A aquestes posicions, els valors al vector de rangs dels valors absoluts de totes les diferències són: 3, 19, 13, 1 i 10 que sumen $R^- = 46$. $R^+ + R^-$ és la suma de tots els rangs, $19(19+1)/2 = 190$, per tant $R^+ = 190 - 46 = 144$ i finalment l'estadístic de test val $W = \min\{144, 46\} = 46$. A les taules del test, per una prova bilateral amb nivell de significació 0.01 i 19 diferències, el valor crític és 32. Com

[3] Com es pot observar, a la columna "Bookstore" hi ha un empat entre els preus de dos llibres. Per simplificar la resolució d'aquesta prova, als llistats hem fet la petita trampa de canviar el preu d'un d'aquests llibres.

que **NO és cert que** 46 sigui menor que 32, no podem rebutjar H_0, **<u>no</u>** hem pogut demostrar que hi hagi diferències de preu en mediana.

2) En les mateixes condicions de l'enunciat, imagina que els estudiants haguessin estudiat més de 2 maneres de comprar els llibres (per exemple, "llibreria", "Internet" i "cooperativa universitària"). Indica el nom d'una prova d'hipòtesis basada en rangs per a intentar demostrar l'existència de diferències de medianes entre els preus segons cada sistema de compra. Indica el nom (o descriu-lo si no en recordes el nom) del disseny experimental sota el qual habitualment aplicaríem aquesta prova.

Una prova adequada seria la de Friedman. És la prova basada en rangs que típicament s'utilitza com a alternativa no paramètrica a l'ANOVA per un disseny en blocs aleatoritzats.

Independentment que existeixin diferències de preu, o no, entre els llibres en llibreria o en línia, un esperaria que hi hagués un cert grau de dependència entre aquests preus. Per exemple, si un llibre és "car" en llibreria també seria d'esperar que ho fos comprat a través d'Internet.

3) Estima el coeficient tau de Kendall entre les variables "Bookstore" i "Online" i explica el significat (la interpretació pràctica, no "significació estadística") del valor obtingut.

Segons els llistats, hi ha n_D = 28 diferències discordants entre totes les diferències possibles, que són 19(19−1)/2 = 171. Per tant, com que no hi ha empats, hi ha n_C = 171 − 28 = 143 diferències concordants i el coeficient de correlació mostral de Kendall ("tau–A") val:

$$\frac{2\left(n_C - n_D\right)}{n\left(n-1\right)} = \frac{2\left(\overbrace{143-28}^{S=115}\right)}{19\left(19-1\right)} = 0.6725$$

Aquest valor és positiu. Vol dir que predominen les concordances, que la tendència general és que si un llibre era més car que un altre a la llibreria, també ho serà per internet, i viceversa.

4) Determina si el coeficient de Kendall és significativament diferent de zero.

Es planteja una prova d'hipòtesis bilateral a la qual la hipòtesi nul·la diu que la tau poblacional és zero i l'alternativa diu que és diferent de zero. Si seguim amb el nivell de significació enunciat al principi, 0.01, a la taula de valors crítics bilaterals de tau per 0.01 i n = 19 hi trobem 0.439. Com que l'estimació obtinguda a l'apartat anterior, 0.6725, és més gran que aquest valor de la taula, podem rebutjar la hipòtesi nul·la i afirmar que la correlació de Kendall poblacional no és zero. Lògicament, s'obtindria el mateix resultat buscant el valor crític per S = 115.

LLISTATS

```
> bookstore = c(132.75, 201.50, 80.00, 153.50, 154.00, 216.00, 199.75, 147.00,
132.00, 182.25, 45.50, 89.95, 123.02, 57.50, 88.25, 189.25, 179.25, 210.25, 66
.75)
> online    = c(136.91, 178.58, 65.00, 120.43, 217.99, 197.10, 168.71, 178.63,
95.89, 145.49, 37.60, 91.69, 148.41, 53.93, 83.69, 133.32, 151.48, 147.30, 55.
16)
>
> d = bookstore - online
> d
 [1]  -4.16  22.92  15.00  33.07 -63.99  18.90  31.04 -31.63  36.11  36.76   7
.90  -1.74 -25.39
[14]   3.57   4.56  55.93  27.77  62.95  11.59
>
> r.bookstore = rank(bookstore)
> r.bookstore
 [1]  9 17  4 11 12 19 16 10  8 14  1  6  7  2  5 15 13 18  3
> r.online = rank(online)
> r.online
 [1] 10 16  4  8 19 18 15 17  7 11  1  6 13  2  5  9 14 12  3
> r.d = rank(d)
> r.d
 [1]  4 12 10 15  1 11 14  2 16 17  8  5  3  6  7 18 13 19  9
> r.abs.d = rank(abs(d))
> r.abs.d
 [1]  3  9  7 14 19  8 12 13 15 16  5  1 10  2  4 17 11 18  6
> r.dades = rank(c(bookstore, online))
> r.dades
 [1] 17 35  8 25 26 37 34 21 16 31  2 11 15  5 10 32 30 36  7 19 28  6 14 38 3
3 27 29 13 20  1 12 23
[33]  3  9 18 24 22  4
>
> x = bookstore
> y = online
>
> n = length(x)
>
> # Taula amb totes les possibles diferències entre x[i] i x[j]:
> difs.x = outer(x,x, "-")
> # Descartem les diferències de la diagonal (i == j) i de la meitat triangula
r superior:
> difs.x = difs.x[ltri <- lower.tri(difs.x)]
> # Totes les possibles diferències entre y[i] i y[j]:
> difs.y = outer(y,y, "-")[ltri]
>
> # No hi ha empats.
> # Diferències que tenen SIGNE DIFERENT, que són discordants, entre x i y:
> sum(difs.x * difs.y < 0)
[1] 28
>
> # Coeficient de correlació de Pearson entre rangs de les columnes de 'preus'
> cor(rank(preus[,1]), rank(preus[,2]))
[1] 0.7982456
```