Capitolo 8. Statistica non-parametrica. Correlazione di Kendall. Correlazione di Spearman. Confronto tra due campioni indipendenti: test di Wilcoxon/Mann-Whitney. Confronto tra due campioni appaiati: test di Wilcoxon. Confronto globale tra più campioni formati da soggetti: test di Kruskal-Wallis. Confronto globale tra più campioni formati dagli stessi soggetti: test di Friedman.

Statistica non-parametrica

Prendiamo lo spunto dal test *t*. Questo test si basa sulla conoscenza della distribuzione del *t* ottenuto da campioni appartenenti alla stessa popolazione (ipotesi nulla). Ma a due condizioni: che i dati siano distribuiti normalmente e i due campioni abbiano varianze uguali (vedi grafico verso la fine del Cap. 2). Infatti, solo conoscendo le proprietà della distribuzione normale, gli statistici hanno avuto modo di valutare la dispersione di certi parametri derivati da questa proponendoli come test per saggiare certe ipotesi. Questa è la storia di tutti i test cosiddetti *parametrici* (la media, la deviazione standard, ecc. sono i parametri statistici). Anche se spesso la natura della distribuzione non è verificata, si *assume* che i dati siano distribuiti normalmente, ed il test è valido a questa condizione.

Sarebbe bene sottoporre sempre i dati ad un test di normalità (es., controllando i parametri globali di skewness, di kurtosis o verificando l'adattamento dell'istogramma alla curva di distribuzione, ma in questo caso occorre che il campione sia di almeno un centinaio di dati).

Quando ci si accorge o solo si sospetta che i dati non siano distribuiti come richiesto dal test (parametrico), si possono fare tre cose:

1. mollare tutto e fare altro (cosa piuttosto facile, ma rinunciataria),
2. trasformare i dati in modo tale da aggiustarne la distribuzione (cosa piuttosto difficile),
3. applicare un appropriato test non-parametrico.

Lasciando perdere le prime due scelte, il test non-parametrico è in grado di condurre ad analoghe, valide conclusioni indipendentemente dai parametri statistici e quindi dal tipo di distribuzione della popolazione.

Molti metodi non-parametrici si basano solo sull'ordine di grandezza dei dati. Come dire la semplice graduatoria dei valori, trascurando i valori stessi. E' in tal modo che la statistica non-parametrica opera: liberandosi dai condizionamenti della distribuzione dei dati, ma al tempo stesso rinunciando ad ottenere da essa ogni possibile informazione.

Dati che non si adattano alla distribuzione normale sono spesso quelli dei punteggi, votazioni, scores, ecc. utilizzati convenzionalmente da un osservatore (il medico, lo psicologo, l'insegnante, il giudice di gara, ecc.) per valutare fenomeni complessi quali l'intelligenza (QI), la capacità di memoria, il rendimento a scuola, la produttività nel lavoro, la prestazione atletica, ecc. In tutti questi casi la scala non è riferita a grandezze fisiche, bensì a diversi livelli qualitativi di espressione del fenomeno, trasformati numericamente solo in base a convenzione (ad es., nei licei si attribuisce 6 per indicare la sufficienza, mentre all'università si attribuisce 18).

Così come la media è un dato squisitamente parametrico, la mediana è l'analogo dato in campo non-parametrico. Mentre la media è *scossa* da valori molto piccoli o molto grandi, la mediana è del tutto stabile rispetto ai cosiddetti *outliers*. Pertanto dovremmo preferire la mediana (e i test non-parametrici) anche quando la distribuzione dei dati sia normale ma sospettiamo la presenza di dati spuri, con valori molto piccoli o molto grandi, dovuti a errori non rilevati nelle procedure sperimentali, nella trascrizione dei dati, ecc.

Esiste tutta una serie di test non-parametrici (v. Siegel, Statistica Non-Parametrica, McGraw-Hill editore) che ripropongono più o meno le stesse ipotesi già viste con i test parametrici. Pertanto, alla domanda: quando applicare i test non-parametrici ? si risponde dicendo che la SNP deve essere applicata quando:

1) i dati non si conformano al tipo di distribuzione richiesto dalle procedure parametriche,
2) i dati si riferiscono a scale ordinali, come quelle associate a punteggi, giudizi, graduatorie, ecc.,
3) tra i dati ve ne sono alcuni molto piccoli o molto grandi (oltre 3 deviazioni standard dalla media: *outliers*) da destare il sospetto che derivino da errore grossolano (es., una trascrizione errata, una dose sbagliata, un preparato confuso con un altro, un tampone inquinato, ecc.) ma che non ci sentiamo di escludere dal campione.

Oltre a tutti questi casi, è possibile ed utile applicare test non-parametrici sempre, anche assieme ai test parametrici, ogni volta che si voglia saggiare una ipotesi prescindendo dalla distribuzione dei dati. C'è da dire che alcuni statistici storcono un po' il naso di fronte a tale considerazione. Certo è che se i campioni sono distribuiti normalmente e differiscono solo per la media, allora il test t è molto più efficiente dell'analogo test non-parametrico di Mann-Whitney. Tuttavia vi sono moltissime situazioni in cui la condizione di normalità è assunta passivamente solo in quanto non si può dimostrare il contrario. Ricordiamo che nei test di normalità (goodness-of-fit test, vedi Cap. 5) la condizione di normalità è data come ipotesi zero, e pertanto essa è da conservare sino a che la probabilità a suo favore non scende sotto il 5%. In tal caso il danno più grave non è dato dai falsi-positivi (ritenere che la distribuzione sia non-normale mentre lo è) ma dai falsi-negativi (ritenere che la distribuzione sia normale mentre non lo è). Oltre a ciò, tutte le volte che consideriamo piccoli campioni con meno di 50 dati, il test non è applicabile. Per cui, nei casi in cui la normalità sia solo presunta, non è affatto male considerare anche il responso del test non-parametrico assieme a quello del test parametrico.

In genere, la SNP converte i dati ponderali nel loro ***rango:*** brutto termine italiano che traduce l'inglese *rank*. L'inglese rank significa infatti anche *posizione in graduatoria/classifica/ordine crescente*, mentre l'italiano rango ha più significato di gerarchia in senso sociale/militare.
I dati:

41	9	84	1	67	123	81

convertiti in rango diventano:

3	2	6	1	4	7	5

Immaginiamo ora alcune varianti:

a) 1230 al posto di 123. I ranghi non cambiano in quanto 1230 ha lo stesso rango di 123:

41	9	84	1	67	**1230**	81

3	2	6	1	4	7	5

b) 12.3 al posto di 123. Alcuni ranghi cambiano di una posizione:

41	9	84	1	67	**12.3**	81

4	2	7	1	5	3	6

c) 0 al posto di 123. Anche in questo caso i ranghi cambiano di poco:

41	9	84	1	67	**0**	81

4	3	7	2	5	1	6

Questi esempi dimostrano come i ranghi siano molto robusti nei confronti di variazioni anche notevoli dei dati. Ma il caso più importante è quello in cui tutti i dati vengono trasformati in modo lineare (es., additivo o moltiplicativo) o non-lineare (es., esponenziale o logaritmico). In tal caso i ranghi non cambiano affatto in quanto i dati mantengono la stessa posizione. In generale, qualsiasi trasformazione, purché monotonica, non altera i ranghi. Come ultimo esempio proponiamo il caso

d) Tutti i dati elevati al quadrato. I ranghi non cambiano:

	41^2	9^2	84^2	1^2	67^2	123^2	81^2
=	**1681**	**81**	**7056**	**1**	**4489**	**15129**	**6561**

3	2	6	1	4	7	5

Ecco perché la SNP prescinde dal tipo di distribuzione. Mentre la distribuzione dei dati è profondamente alterata da trasformazioni non-lineari (la trasformazione esponenziale stira la coda a destra, mentre quella logaritmica stira la coda a sinistra) i dati mantengono gli stessi ranghi.
Attenzione: le cose non filano lisce se esistono dati negativi. Infatti in tal caso i quadrati dei valori negativi si rifletterebbero sulla scala dei valori positivi sconvolgendo completamente l'ordine originario (occorre prestare molta attenzione nell'applicare trasformazioni in presenza di valori negativi).

Nota bene: quando esistono valori uguali, a ciascuno di essi si attribuisce la media dei ranghi che spetterebbero agli stessi valori se questi fossero diversi.

Ad esempio i ranghi della serie:

32	63	41	85	32	51	85	79	85	27	68

sono:

2.5	6	4	**10**	**2.5**	5	**10**	8	**10**	1	7

Ne deriva che, anche in presenza di sovrapposizioni, la **somma dei ranghi** di una serie di **n** dati è comunque:

$$\sum r = \frac{n(n+1)}{2}$$

Pertanto, il **rango medio** di una serie di **n** dati sarà:

$$r_{medio} = \frac{n(n+1)}{2}\frac{1}{n} = \frac{n+1}{2}$$

Infine, non bisogna credere che la SNP comporti sempre la trasformazione in rango. Anche se non menzionato a suo tempo, il test del χ^2 è anch'esso un test non-parametrico per dati di frequenze. Se infatti consideriamo ad esempio la semplice tabella 2×2, notiamo che l'ipotesi nulla non fa riferimento a nessun tipo di distribuzione in particolare, ma solo ai rapporti dei dati all'interno della stessa tabella ed al contrasto risultante tra le frequenze osservate e le frequenze attese ottenute ipotizzando uguali proporzioni tra colonne (o tra righe, che è lo stesso).

Di seguito sono illustrati alcuni test non-parametrici. Data l'analogia con alcuni test parametrici già visti e la semplicità di applicazione, l'esposizione è condotta a maniera di esercizio.

Correlazione di Kendall

step 1.
I dati Y vengono riscritti ordinati a seconda del valore crescente della variabile X .
step 2.
Si considera la successione dei valori della variabile Y. Per ogni valore di Y si conta il numero di successivi valori maggiori ($S_{maggiori}$) e il numero di successivi valori minori (S_{minori}). Se tra i successivi si incontra un valore uguale si incrementa di 0.5 sia il totale $S_{maggiori}$ sia il totale S_{minori}.
step 3.
La somma di $S_{maggiori}$ + S_{minori} deve dare n(n-1)/2, altrimenti i conti sono sbagliati.
step 4.
Si calcola la differenza $S_{diff} = S_{maggiori} - S_{minori}$.
Si confronta il valore di S_{diff} con i valori critici riportati in tabella per verificare la significatività della correlazione.

Quando Y varia in modo del tutto irregolare rispetto a X avremo $S_{maggiori} = S_{minori}$, per cui S_{diff} sarà **nullo**: $S_{diff} = 0$
Quando Y cresce costantemente al crescere di X, avremo $S_{minori}=0$, per cui S_{diff} sarà **massimo**: $S_{diff} = S_{maggiori} = n(n-1)/2$
Viceversa, quando Y decresce costantemente al crescere di X, avremo $S_{maggiori}=0$, per cui S_{diff} sarà **minimo**: $S_{diff} = S_{minori} = -n(n-1)/2$

step 5.
La definizione del range di S_{diff} entro i suoi valori minimo e massimo consente la valutazione di un coefficiente di correlazione non-parametrico **t** (tàu di Kendall) che oscilla tra +1 e -1:

$$t = \frac{S_{diff}}{\frac{n(n-1)}{2}}$$

Esempio:

dati iniziali

X	Y
8	11
13	14
6.6	5.9
3	4.5
9	10
11	15
5	2

step 1. *step 2.*

X	Y	considerando la sequenza dei valori Y dall'alto verso il basso ordinata in funzione dei valori crescenti di X calcolare per ciascun valore di Y...	
		da quanti valori maggiori è seguito	da quanti valori minori è seguito
3	4.5	5	1
5	2	5	0
6.6	5.9	4	0
8	11	2	1
9	10	2	0
11	15	0	1
13	14	-	-
		totale $S_{maggiori}=18$	totale $S_{minori}=3$

step 3.
Verifica calcolo dei ranghi e delle somme:
$n(n-1)/2=21$
$S_{maggiori} + S_{minori} = 18+3 = 21$

step 4.
$S_{diff}= 18-3 = 15$

step 5.
$\tau=15/21=0.71$

Correlazione di Spearman

step 1.
I dati vengono trasformati nei loro ranghi, con ordinamenti diversi per le due variabili X e Y.
step 2.
Sui ranghi, si valuta la normale correlazione utilizzata dai metodi parametrici attraverso il calcolo della codevianza e delle devianze di X e Y. Notare che le due devianze sono uguali, poiché X e Y hanno uguali valori.
Il coefficiente di correlazione non-parametrico di Spearman oscilla anch'esso tra +1 e -1, anche se non corrisponde necessariamente al τ di Kendall.

Esempio:

dati iniziali

X	Y
8	11
13	14
6.6	5.9
3	4.5
9	10
11	15
5	2

step 1.

X	Y
4	5
7	6
3	3
1	2
5	4
6	7
2	1

step 2.
Codevianza=25

$\text{Devianza}_x = \text{Devianza}_y = 28$

$$r = \frac{25}{\sqrt{28 \cdot 28}} = \frac{25}{28} = 0.89$$

Test di Wilcoxon / Mann-Whitney per il confronto tra due campioni indipendenti o test della somma dei ranghi
E' l'analogo non-parametrico del test *t* di Student per campioni indipendenti.

step 1.
Ordinare i dati in rango, comprendendo nello stesso ordinamento i due campioni. Se i campioni non sono bilanciati, diciamo che il campione più piccolo ha numerosità **n** e quello più grande ha numerosità **m**.
La somma dei ranghi dei due campioni è

$$\Sigma_r = \frac{(n+m)(n+m+1)}{2}$$

l'ipotesi nulla di assortimento casuale dei valori nei due gruppi prevede che entrambe i gruppi abbiano lo stesso rango medio, e cioè

$$r_m = \frac{(n+m)(n+m+1)}{2}\frac{1}{(n+m)} = \frac{n+m+1}{2}$$

Moltiplicando il rango medio per le rispettive numerosità, si ricava la somma dei ranghi attesa per la condizione di indifferenza.

step 2.
Sommare i ranghi del campione più piccolo (**T**). Tale somma è già il risultato del test. Infatti quanto più la somma dei ranghi del campione più piccolo si discosta da quella attesa, tanto più la probabilità che i due campioni siano assortimenti casuali di ranghi perde consistenza. La verifica in tabella ci dirà se mantenere l'ipotesi nulla o rifiutarla oltre le diverse soglie critiche di significatività. Per ogni livello di significatività, in tabella sono tabulati due valori estremi: uno molto piccolo ed uno molto grande. La significatività si raggiungerà se **T** sarà minore del valore più piccolo o maggiore del valore più grande in tabella.
Si potrebbe considerare la somma dei ranghi del campione più grande - che è lo stesso, in quanto le due somme sono mutualmente vincolate. Ci si basa sul campione più piccolo per fare economia di tabelle.
Nota: vi è una versione differente del test nota come Test U di Mann-Whitney. I risultati sono comunque gli stessi.

step 3.
Comunque, per verificare che abbiamo fatto bene i conti verifichiamo che le somme dei ranghi dei due campioni dìano insieme il valore totale.

step 4.
Se i campioni sono sufficientemente grandi (m>8) è possibile fare a meno della tabella e sfruttare il fatto che **T** tende a distribuirsi normalmente attorno al valore atteso dall'ipotesi nulla

$$m_T = \text{numerosità} \times \text{rango medio} = n \cdot \frac{n+m+1}{2}$$

con deviazione standard pari a

$$s_T = \sqrt{\frac{n \cdot m \cdot (n+m+1)}{12}}$$

per cui si può calcolare la deviata standardizzata:

$$z_T = \frac{T - m_T}{s_T} = \frac{T - \left(n \cdot \frac{n+m+1}{2}\right)}{\sqrt{\frac{n \cdot m \cdot (n+m+1)}{12}}}$$

Se z_T supera in valore assoluto il solito 1.96 i due gruppi possono considerarsi significativamente differenti. La suddetta formula andrebbe perfezionata (al numeratore) con la correzione per la continuità e (al denominatore) con la correzione della deviazione standard per la presenza di eventuali valori sovrapposti. Per questo si rimanda a manuali più completi.

Esempio:

dati iniziali

A	B
7.2	1
10	4
12	5
15	6
20	8
23	8
41	8
60	10
	11
	12
	15
	21
n=8	m=12

step1.

A	B
5	1
9.5	2
12.5	3
14.5	4
16	7
18	7
19	7
20	9.5
	11
	12.5
	14.5
	17
step 2.	
T_A=114.5	T_B=95.5

step 3.

$$114.5 + 95.5 = 210$$

$$\frac{(n+m)(n+m+1)}{2} = \frac{(8+12)(8+12+1)}{2} = \frac{20 \cdot 21}{2} = 210$$

OK

step 4.

$$z_T = \frac{114.5 - \left(8 \cdot \frac{8+12+1}{2}\right)}{\sqrt{\frac{8 \cdot 12 \cdot (8+12+1)}{12}}} = \frac{30.5}{12.96} = 2.35$$

Test di Wilcoxon per il confronto tra due campioni appaiati o test dei segni
E' l'analogo non-parametrico del test *t* di Student per campioni appaiati.

step 1.
Calcolare le differenze tra i due campioni.

step 2.
Assegnare i ranghi alle differenze trascurando i segni (cioè i ranghi ai valori assoluti delle differenze).

step 3.
Applicare i segni delle differenze ai ranghi.

step 4.
Calcolare la somma dei ranghi 'segnati' (**W**). Con tale valore si entra in tabella. Notare che in alcune versioni del test si calcolano separatamente le somme dei ranghi positivi e dei ranghi negativi e con la minore delle due si entra in tabella - ovviamente una diversa tabella.

E' chiaro che se i valori dei due campioni sono casualmente assortiti, le differenze si annullano tra loro e così anche la somma dei ranghi tende a zero.
Viceversa, se vi è una variazione coerente tra i due campioni, le differenze tendono ad essere tutte positive o tutte negative, per cui la somma totale dei ranghi 'segnati' tende a crescere sino al limite del valore assoluto massimo

$$W_{max} = \frac{n(n+1)}{2}$$

Se n>20, si può fare a meno della tabella e sfruttare il fatto che la distribuzione di W tende a diventare normale con media

m_W=0

e deviazione standard

$$s_W = \sqrt{\frac{n(n+1)(2n+1)}{6}}$$

step 5.
Per cui il test si può proporre come il calcolo della deviata standardizzata

$$z_W = \frac{W-0}{\sqrt{\frac{n(n+1)(2n+1)}{6}}}$$

Ovviamente, se z_W supera in valore assoluto il solito 1.96 la differenza tra i due gruppi può considerarsi significativa.

Come per il precedente caso, vi è la possibilità di introdurre la correzione per la continuità e per la sovrapposizione di dati (cosa che qui non prendiamo in considerazione).

Esempio:

dati iniziali *step1.* *step 2.* *step 3.*

prima	dopo	differenza	valore assoluto	rango	rango segnato
3	7	4	4	6	6
5	4	-1	1	2	-2
2	9	7	7	8	8
11	16	5	5	7	7
6	5	-1	1	2	-2
7	10	3	3	5	5
4	6	2	2	4	4
2	1	-1	1	2	-2
					step 4.
					W=24

step 5.

$$z_W = \frac{24}{\sqrt{\frac{8(8+1)(16+1)}{6}}} = 1.68$$

Test di Kruskal-Wallis per il confronto globale di più gruppi
E' l'analogo non-parametrico dell'analisi della varianza.

step 1.
Assegnare i ranghi ai dati, considerando tutti i dati dei vari gruppi messi assieme (come per Wilcoxon/Mann-Whitney)

step 2.
Calcolare la somma dei ranghi e poi il rango medio per ciascun gruppo r_{mgr}.

step 3.
Calcolare l'esattezza dei ranghi confrontando la somma delle somme dei gruppi con N(N+1)/2, ove N è il numero totale di dati.

step 4.
Calcolare il rango medio generale di tutti i dati secondo la solita formula: r_{mgen}= (N+1)/2.

step 5.
Calcolare la differenza tra il rango medio di ciascun gruppo (r_{mgr}) ed il rango medio generale (r_{mgen}), elevare al quadrato e moltiplicare per la numerosità del gruppo (n_{gr}). Quindi fare la sommatoria:

$$D = \sum n_{gr} (r_{mgr} - r_{m\,gen})^2$$

Il procedimento ha forti analogie col calcolo della devianza tra gruppi per l'analisi della varianza.

step 6.
Normalizzare D, calcolando la statistica:

$$H = \frac{D}{\frac{N(N+1)}{12}}$$

H è la statistica da confrontare con i valori tabulati per verificare la significatività delle differenze tra i gruppi.

Se i gruppi sono abbastanza numerosi (con oltre 3 gruppi e 5 soggetti per gruppo) la distribuzione di H è approssimabile a quella del χ^2 con tanti gradi di libertà quanti sono i gruppi meno 1.

Esempio:

dati iniziali.

Gruppi				
A	B	C	D	E
10	26	40	24	35
14	30	34	11	29
18	27	23	17	37
20	12	36	21	13
22	15	33	19	28
16	31	38	32	39
		41		42
				43
n_A=6	n_B=6	n_C=7	n_D=6	n_E=8
N=6+6+7+6+8=33				

step 1.

A	B	C	D	E
1	16	30	15	25
5	20	24	2	19
9	17	14	8	27
11	3	26	12	4
13	6	23	10	18
7	21	28	22	29
		31		32
				33
step 2.				
Σr_A=46	Σr_B=83	Σr_C=176	Σr_D=69	Σr_E=187
r_{mA}=7.7	r_{mB}=13.8	r_{mC}=25	r_{mD}=11.5	r_{mE}=23.3

step 3.
46+83+176+69+187=561
N(N+1)/2=561

OK

step 4.

r_{mgen}= (N+1)/2=(33+1)/2=17

step 5.

$D=6(7.7 - 17)^2 + 6(13.8 - 17)^2 + 7(25 - 17)^2 + 6(11.5 - 17)^2 + 8(23.3 - 17)^2 = 1527.4$

step 6.

$$H = \frac{1527.4}{\frac{33(33+1)}{12}} = 16.34$$

Test di Friedman per il confronto di più trattamenti applicati agli stessi soggetti

E' l'analogo non-parametrico dell'analisi della varianza cosiddetta a due vie.
Un gruppo di n soggetti, sempre gli stessi, è sottoposto a k diversi trattamenti. Ad es., un gruppo di operai posti a lavorare in ambienti con diverse condizioni di luce, rumore, spazio, ecc.; oppure un gruppo di pazienti che –in diverse occasioni – ricevono terapie differenziate, ecc.

n = numero dei soggetti
k = numero dei trattamenti

step 1.
Assegnare i ranghi ai dati di ciascun soggetto, valutandoli separatamente nell'ambito di ciascun soggetto (operando quindi riga per riga).

step 2.
Calcolare il rango medio di ciascun trattamento r_{mtrat} (operando quindi colonna per colonna).

step 3.
Calcolare il rango medio atteso di ciascun trattamento nell'ipotesi in cui tutti i dati siano uguali (ipotesi di indifferenza)

$$r_{m\ atteso} = \frac{(k+1)}{2}$$ (uguale per tutti i trattamenti)

step 4.
Sommare quindi, per ciascun trattamento, il quadrato della differenza tra rango medio osservato e rango medio atteso moltiplicata per il numero di soggetti:

$$S = \sum\left[n(r_{mtrat} - r_{matteso})^2\right]$$

step 5.
Normalizzare S con il rapporto:

$$c_r^2 = \frac{S}{\frac{nk(k+1)}{12}}$$

χ_r^2 è da confrontare con i valori tabulati per verificare la significatività delle differenze tra i campioni. Quando i dati sono abbastanza numerosi (almeno 9 soggetti e 3 trattamenti, ecc.), χ_r^2 tende a distribuirsi come χ^2 con un numero di gradi di libertà pari al numero di campioni meno 1.

Nota: così come il test di Kruskal-Wallis mostra una stretta analogia con l'ANOVA parametrica ad una via (o ad un criterio di classificazione), il test di Friedman è analogo all'ANOVA parametrica a due vie, tenendo presente tuttavia che Friedman valuta solo le differenze tra i campioni, e non si esprime su possibili differenze tra i soggetti, cosa che invece l'ANOVA a due vie fa.

Esempio:

dati iniziali.

	trattamenti			
soggetti	A	B	C	D
a	21	65	39	10
b	43	87	61	6
c	32	43	83	29
d	15	76	58	4
e	17	43	34	10
f	34	49	54	29

step 2.

	trattamenti			
soggetti	A	B	C	D
a	2	4	3	1
b	2	4	3	1
c	2	3	4	1
d	2	4	3	1
e	2	4	3	1
f	2	3	4	1
	r_{mA} =12/6 =2	r_{mB} =22/6 =3.667	r_{mC} =20/6 =3.333	r_{mD} =6/6 =1

step 3.

$$r_{matteso} = \frac{(4+1)}{2} = 2.5$$

step 4.

$$S = 6^2(2-2.5)^2 + 6^2(3.667-2.5)^2 + 6^2(3.333-2.5)^2 + 6^2(1-2.5)^2 = 164$$

step 5.

$$c_r^2 = \frac{164}{\frac{6 \cdot 4(4+1)}{12}} = \frac{164}{10} = 16.4$$