Esercizio 1

Si supponga di aver assegnato ad una popolazione di N = 4 dattilografe un test e di aver ottenuto i seguenti risultati:

Dattilografa	N. Errori
A	3
B	2
C	1
D	4

La variabile X, il numero di errori commessi da ciascuna dattilografa, ha $\mu_X = 2,5$ e $\sigma_X^2 = 1,25$.
Determinare l'universo campionario rispettivamente per n = 2, n = 3 e n = 4 immaginando di procedere con un campionamento casuale semplice (con reintroduzione).
Inoltre, per i differenti valori di n determinare:
1) la distribuzione della variabile casuale media campionaria, calcolandone il valore atteso e la varianza;
2) la distribuzione della variabile casuale varianza campionaria, calcolandone il valore atteso.

Svolgimento

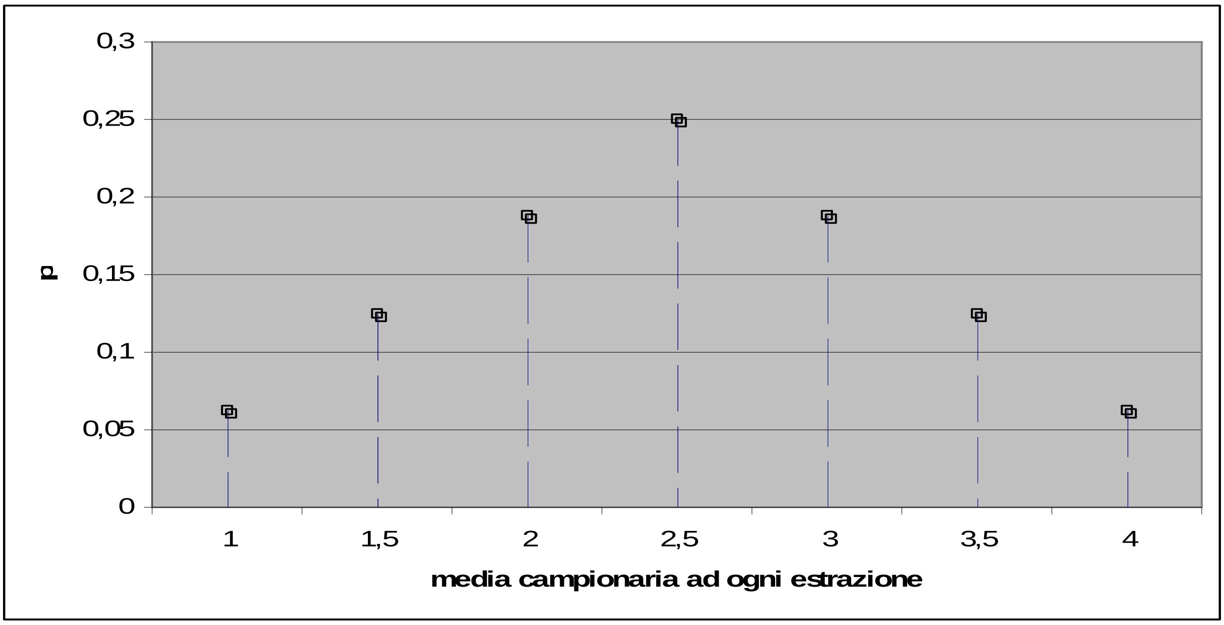
Trattandosi di campionamento con reintroduzione, l'universo campionario è pari a N^n.
Per n = 2, quindi, i possibili campioni sono $4^2 = 16$, e cioè:

Campione	Dattilografe	Risultati del campione	Media Campionaria
1	A,A	3,3	3,0
2	A,B	3,2	2,5
3	A,C	3,1	2,0
4	A,D	3,4	3,5
5	B,A	2,3	2,5
6	B,B	2,2	2,0
7	B,C	2,1	1,5
8	B,D	2,4	3,0
9	C,A	1,3	2,0
10	C,B	1,2	1,5
11	C,C	1,1	1,0
12	C,D	1,4	2,5
13	D,A	4,3	3,5
14	D,B	4,2	3,0
15	D,C	4,1	2,5
16	D,D	4,4	4,0

Supponendo che ogni dattilografa ha la medesima probabilità di essere estratta nel campione e ciò che interessa è la media campionaria, cioè il numero medio di errori commessi da ogni campione, la funzione di probabilità è la seguente:

$P(x_i)$	0,0625	0,125	0,1875	0,25	0,1875	0,125	0,0625
$\bar{X}_i$	1	1,5	2	2,5	3	3,5	4

Come si può facilmente calcolare, $E(\bar{X}) = 2,5$ e $\mathrm{Var}(\bar{X}) = 0,625$.



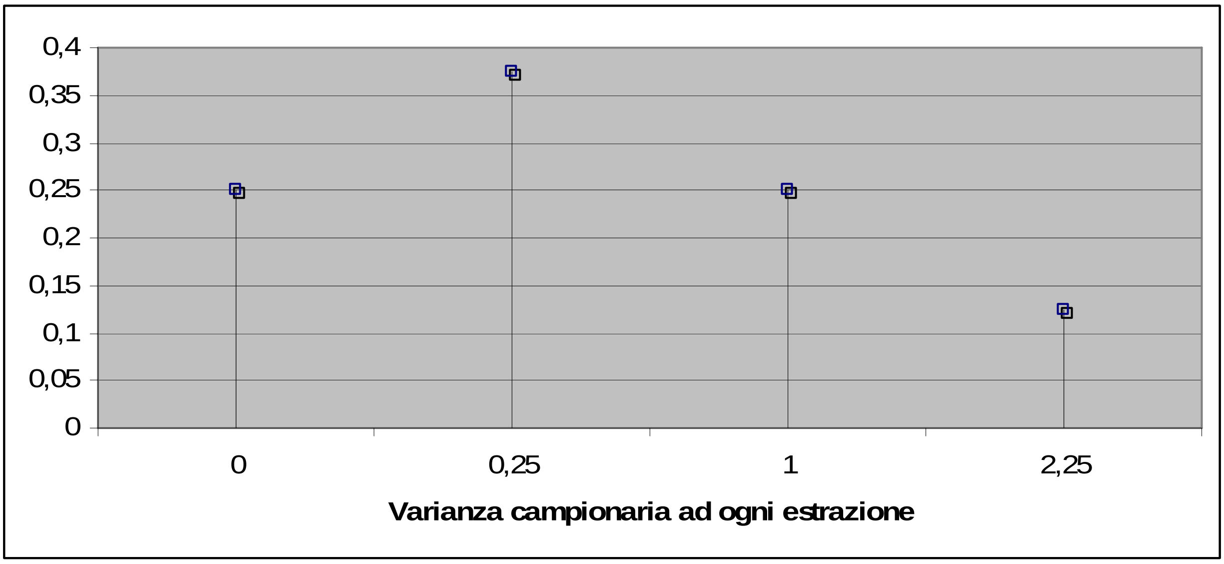
Per determinare la distribuzione della variabile casuale varianza campionaria bisogna calcolare la statistica varianza campionaria $\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2$ ad ogni estrazione:

Campione	Dattilografe	Risultati del campione	Media Campionaria	Varianza Campionaria
1	A,A	3,3	3,0	0,00
2	A,B	3,2	2,5	0,25
3	A,C	3,1	2,0	1,00
4	A,D	3,4	3,5	0,25
5	B,A	2,3	2,5	0,25
6	B,B	2,2	2,0	0,00
7	B,C	2,1	1,5	0,25
8	B,D	2,4	3,0	1,00
9	C,A	1,3	2,0	1,00
10	C,B	1,2	1,5	0,25
11	C,C	1,1	1,0	0,00
12	C,D	1,4	2,5	2,25
13	D,A	4,3	3,5	0,25
14	D,B	4,2	3,0	1,00
15	D,C	4,1	2,5	2,25
16	D,D	4,4	4,0	0,00

la funzione di probabilità è la seguente:

$P(x_i)$	0,25	0,375	0,25	0,125
$Var(X_i)$	0	0,25	1	2,25

Come si può facilmente calcolare, $E(Var(X_i)) = 0,625$



***

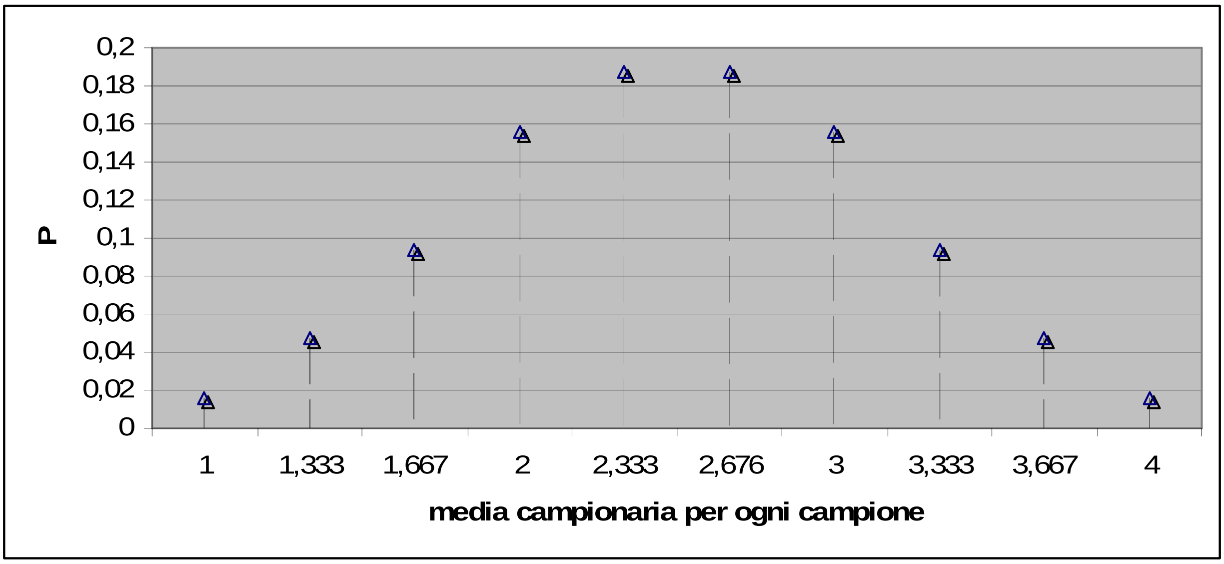
Si ripete l'esercizio per n = 3, dove l'universo campionario è pari a $4^3 = 64$, e cioè:

Campione	Dattilografe	Risultati del campione	Media Campionaria	Campione	Dattilografe	Risultati del campione	Media Campionaria
1	A,A,A	3,3,3	3,000	33	C,A,A	1,3,3	2,333
2	A,A,B	3,3,2	2,667	34	C,A,B	1,3,2	2,000
3	A,A,C	3,3,1	2,333	35	C,A,C	1,3,1	1,667
4	A,A,D	3,3,4	3,333	36	C,A,D	1,3,4	2,667
5	A,B,A	3,2,3	2,667	37	C,B,A	1,2,3	2,000
6	A,B,B	3,2,2	2,333	38	C,B,B	1,2,2	1,667
7	A,B,C	3,2,1	2,000	39	C,B,C	1,2,1	1,333
8	A,B,D	3,2,4	3,000	40	C,B,D	1,2,4	2,333
9	A,C,A	3,1,3	2,333	41	C,C,A	1,1,3	1,667
10	A,C,B	3,1,2	2,000	42	C,C,B	1,1,2	1,333
11	A,C,C	3,1,1	1,667	43	C,C,C	1,1,1	1,000
12	A,C,D	3,1,4	2,667	44	C,C,D	1,1,4	2,000
13	A,D,A	3,4,3	3,333	45	C.D.A	1,4,3	2,667
14	A.D.B	3,4,2	3,000	46	C.D.B	1,4,2	2,333
15	A,D,C	3,4,1	2,667	47	C,D,C	1,4,1	2,000
16	A,D,D	3,4,4	3,667	48	C,D,D	1,4,4	3,000
17	B,A,A	2,3,3	2,667	49	D,A,A	4,3,3	3,333
18	B,A,B	2,3,2	2,333	50	D,A,B	4,3,2	3,000
19	B,A,C	2,3,1	2,000	51	D,A,C	4,3,1	2,667
20	B,A,D	2,3,4	3,000	52	D,A,D	4,3,4	3,667
21	B,B,A	2,2,3	2,333	53	D,B,A	4,2,3	3,000
22	B,B,B	2,2,2	2,000	54	D,B,B	4,2,2	2,667
23	B,B,C	2,2,1	1,667	55	D,B,C	4,2,1	2,333
24	B,B,D	2,2,4	2,667	56	D,B,D	4,2,4	3,333
25	B,C,A	2,1,3	2,000	57	D,C,A	4,1,3	2,667
26	B,C,B	2,1,2	1,667	58	D,C,B	4,1,2	2,333
27	B,C,C	2,1,1	1,333	59	D,C,C	4,1,1	2,000
28	B,C,D	2,1,4	2,333	60	D,C,D	4,1,4	3,000
29	B,D,A	2,4,3	3,000	61	D,D,A	4,4,3	3,667
30	B,D,B	2,4,2	2,667	62	D,D,B	4,4,2	3,333
31	B,D,C	2,4,1	2,333	63	D,D,C	4,4,1	3,000
32	B,D,D	2,4,4	3,333	64	D,D,D	4,4,4	4,000

Supponendo che ogni dattilografa ha la medesima probabilità di essere estratta nel campione e ciò che interessa è la media campionaria, cioè il numero medio di errori commessi da ogni campione, la funzione di probabilità è la seguente:

$P(x_i)$	0,015625	0,046875	0,09375	0,15625	0,1875	0,1875	0,15625	0,09375	0,046875	0,015625
$\bar{X}_i$	1	1,333	1,667	2	2,333	2,667	3	3,333	3,667	4

Come si può facilmente calcolare, $E(\bar{X}) = 2{,}5$ e Var $(\bar{X}) = 0{,}416667$.


0,2
0,18
0,16
0,14
0,12
0,1
0,08
0,06
0,04
0,02
0
P
1
1,333
1,667
2
2,333
2,676
3
3,333
3,667
4
media campionaria per ogni campione

Per determinare la distribuzione della variabile casuale varianza campionaria bisogna calcolare la statistica varianza campionaria $\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2$ ad ogni estrazione:

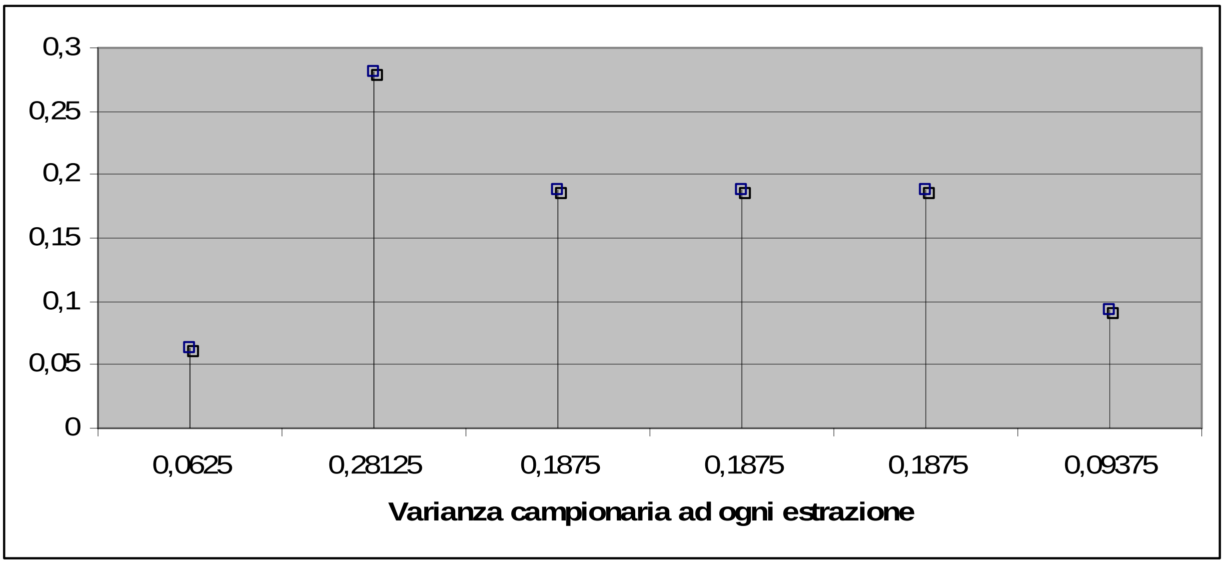
Camp.	Datt.	Ris	$\bar{X}$	Var(X)
1	A,A,A	3,3,3	3,000	0,000
2	A,A,B	3,3,2	2,667	0,222
3	A,A,C	3,3,1	2,333	0,889
4	A,A,D	3,3,4	3,333	0,222
5	A,B,A	3,2,3	2,667	0,222
6	A,B,B	3,2,2	2,333	0,222
7	A,B,C	3,2,1	2,000	0,667
8	A,B,D	3,2,4	3,000	0,667
9	A,C,A	3,1,3	2,333	0,889
10	A,C,B	3,1,2	2,000	0,667
11	A,C,C	3,1,1	1,667	0,889
12	A,C,D	3,1,4	2,667	1,556
13	A,D,A	3,4,3	3,333	0,222
14	A.D.B	3,4,2	3,000	0,667
15	A,D,C	3,4,1	2,667	1,556
16	A,D,D	3,4,4	3,667	0,222
17	B,A,A	2,3,3	2,667	0,222
18	B,A,B	2,3,2	2,333	0,222
19	B,A,C	2,3,1	2,000	0,667
20	B,A,D	2,3,4	3,000	0,667
21	B,B,A	2,2,3	2,333	0,222
22	B,B,B	2,2,2	2,000	0,000
23	B,B,C	2,2,1	1,667	0,222
24	B,B,D	2,2,4	2,667	0,889
25	B,C,A	2,1,3	2,000	0,667
26	B,C,B	2,1,2	1,667	0,222
27	B,C,C	2,1,1	1,333	0,222
28	B,C,D	2,1,4	2,333	1,556
29	B,D,A	2,4,3	3,000	0,667
30	B,D,B	2,4,2	2,667	0,889
31	B,D,C	2,4,1	2,333	1,556
32	B,D,D	2,4,4	3,333	0,889

Camp.	Datt.	Ris	$\bar{X}$	Var(X)
33	C,A,A	1,3,3	2,333	0,889
34	C,A,B	1,3,2	2,000	0,667
35	C,A,C	1,3,1	1,667	0,889
36	C,A,D	1,3,4	2,667	1,556
37	C,B,A	1,2,3	2,000	0,667
38	C,B,B	1,2,2	1,667	0,222
39	C,B,C	1,2,1	1,333	0,222
40	C,B,D	1,2,4	2,333	1,556
41	C,C,A	1,1,3	1,667	0,889
42	C,C,B	1,1,2	1,333	0,222
43	C,C,C	1,1,1	1,000	0,000
44	C,C,D	1,1,4	2,000	2,000
45	C.D.A	1,4,3	2,667	1,556
46	C.D.B	1,4,2	2,333	1,556
47	C,D,C	1,4,1	2,000	2,000
48	C,D,D	1,4,4	3,000	2,000
49	D,A,A	4,3,3	3,333	0,222
50	D,A,B	4,3,2	3,000	0,667
51	D,A,C	4,3,1	2,667	1,556
52	D,A,D	4,3,4	3,667	0,222
53	D,B,A	4,2,3	3,000	0,667
54	D,B,B	4,2,2	2,667	0,889
55	D,B,C	4,2,1	2,333	1,556
56	D,B,D	4,2,4	3,333	0,889
57	D,C,A	4,1,3	2,667	1,556
58	D,C,B	4,1,2	2,333	1,556
59	D,C,C	4,1,1	2,000	2,000
60	D,C,D	4,1,4	3,000	2,000
61	D,D,A	4,4,3	3,667	0,222
62	D,D,B	4,4,2	3,333	0,889
63	D,D,C	4,4,1	3,000	2,000
64	D,D,D	4,4,4	4,000	0,000

la funzione di probabilità è la seguente:

$P(x_i)$	0,0625	0,28125	0,1875	0,1875	0,1875	0,09375
$Var(X_i)$	0	0,222222	0,666667	0,888889	1,555556	2

Come si può facilmente calcolare, $E(Var(X_i)) = 0,833\overline{3}$



***

Infine, per n = 4 l'universo campionario è pari a $4^4 = 256$.

La tabella di tutti i possibili campioni sarebbe troppo grande da raffigurare, così come il calcolo della media campionaria e della varianza campionaria.
Però, come è noto e come è stato possibile verificare empiricamente nei precedenti passaggi, sappiamo che $E(\overline{X}) = \mu_X$ e che $Var(\overline{X}) = \dfrac{\sigma_X^2}{n}$.
Infatti:

$$E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = E\left(\frac{1}{n} S_n\right) = \frac{1}{n} E\left(S_n\right) = \frac{1}{n} n\mu = \mu$$

$$Var\left(\overline{X}\right) = Var\left(\frac{1}{n} S_n\right) = \frac{1}{n^2} Var\left(S_n\right) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

Quindi, per n = 4, $E\left(\overline{X}\right) = 2,5$ $Var\left(\overline{X}\right) = \dfrac{1,25}{4} = 0,3125$

ALL'AUMENTARE DELLA NUMEROSITA' CAMPIONARIA LA DISPERSIONE INTORNO AL VALORE ATTESO DIMINUISCE

Per quanto riguarda la V.C. varianza campionaria, si è potuto constatare che il suo valore atteso è diverso dal parametro (in questo caso supposto) incognito della popolazione, in altri termini è uno stimatore non corretto (o distorto).

Tuttavia il fattore di correzione è noto, ed è pari a $\frac{n}{n-1}$.

Infatti,

$E\left(\tilde{S}^2\right) = \frac{1}{n}\sum_i E\left(X_i - \overline{X}\right)^2$, aggiungendo e sottraendo la costante μ si ottiene:

$$E\left(\tilde{S}^2\right) = \frac{1}{n}E\sum_i\left[\left(X_i - \mu\right) - \left(\overline{X} - \mu\right)\right]^2 = \frac{1}{n}E\left[\sum_i\left(X_i - \mu\right)^2 - n\left(\overline{X} - \mu\right)^2\right] =$$

$$= \frac{1}{n}\left(n\sigma^2 - n\frac{\sigma^2}{n}\right) = \sigma^2 - \frac{\sigma^2}{n} = \sigma^2\left(\frac{n-1}{n}\right)$$

Quindi:

per n=2 si ha $\tilde{S}^2 = 0,625$ da cui $\sigma^2 = 0,625\frac{2}{1} = 1,25$;

per n=3 si ha $\tilde{S}^2 = 0,833\overline{3}$ da cui $\sigma^2 = 0,833\overline{3}\frac{3}{2} = 1,25$;

per n=4 si ha $\tilde{S}^2 = \sigma^2\frac{n-1}{n} = 1,25\frac{3}{4} = 0,9375$