Domande multiple choice

Per evitare la distorsione della stima occorre (sono possibili più alternative)?

- ◯ Aumentare la dimensione del campione
- ◯ Controllare l'adeguatezza dello strumento di rilevazione
- ◯ Controllare il metodo di selezione del campione
- ◯ Verificare il rispetto delle direttive del protocollo di rilevazione

Nel campionamento casuale stratificato il campione è composto da tanti campioni casuali semplici quanti sono gli strati?

- ◯ SI
- ◯ NO

In riferimento al bias di selezione quali delle seguenti affermazioni non è corretta?

- ◯ Compromette la validità di un esperimento
- ◯ E' un errore sistematico
- ◯ E' un errore casuale

Quali delle seguenti affermazione è vera (sono possibili più alternative)?

- ◯ Nel rispetto del requisito di simultaneità, il censimento viene condotto nell'arco di un periodo di tempo limitato
- ◯ Dati più dettagliati possano essere ottenuti attraverso il ricorso a tecniche campionarie
- ◯ L'indagine sulle forze di lavoro utilizza un campione non probabilistico a 2 stadi
- ◯ Nella quantificazione dei costi gioca un ruolo fondamentale anche la tecnica di rilevazione adottata (es tipo di intervista)

L'estrazione di un campione da un elenco telefonico è un campione probabilistico?

- ◯ SI
- ◯ NO

Con riferimento ai campioni probabilistici quale delle seguenti affermazioni è falsa?

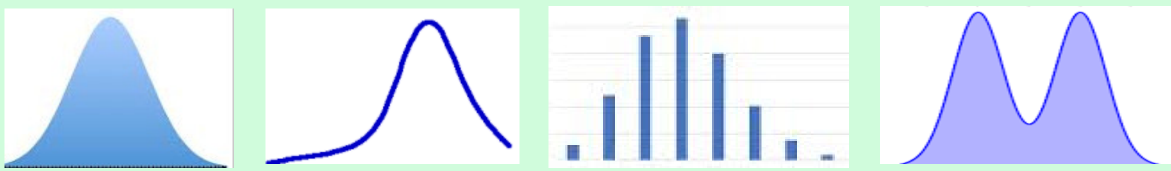
- ◯ Le unità statistiche sono estratte a sorte
- ◯ Eventuali scarti devono essere non significativi ed imputabili al caso
- ◯ Le unità statistiche da includere nel campione sono scelte in modo ragionato
- ◯ La scelta del campione non risulta influenzata dal ricercatore stesso

Teorema del limite centrale

Il **teorema del limite centrale** afferma che quando l'ampiezza del campione casuale diventa sufficientemente grande (n > 30), la distribuzione delle medie campionarie può essere approssimata dalla distribuzione di probabilità normale con media μ e deviazione standard $\sigma/\sqrt{n}$.

$$\overline{X} \sim N\left(\mu, \sigma/\sqrt{n}\right)$$

Ciò è possibile indipendentemente dalla forma della distribuzione dei singoli valori della popolazione.

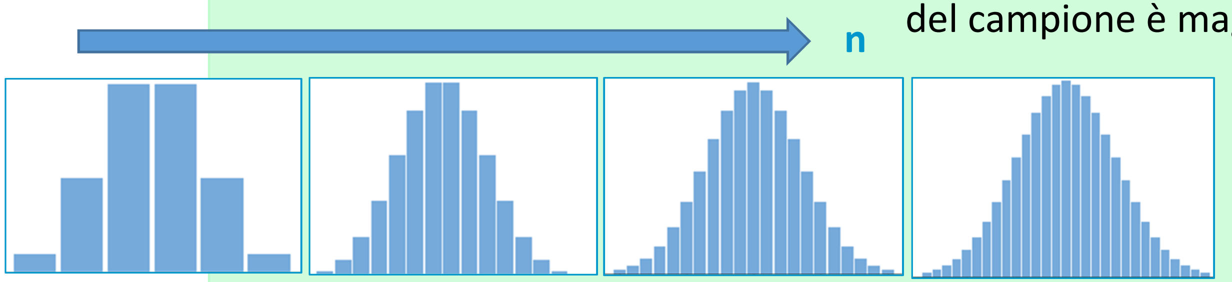


Teorema del limite centrale

Se la popolazione d'interesse è distribuita come una normale → La distribuzione campionaria è distribuita come una normale indipendentemente dall'ampiezza del campione

$N\left(\mu, \sigma/\sqrt{n}\right)$

Se la popolazione non è distribuita come una normale → La distribuzione campionaria è distribuita **approssimativamente** come una normale, se l'ampiezza del campione è maggiore di 30.



$\sim N\left(\mu, \sigma/\sqrt{n}\right)$

Esempio

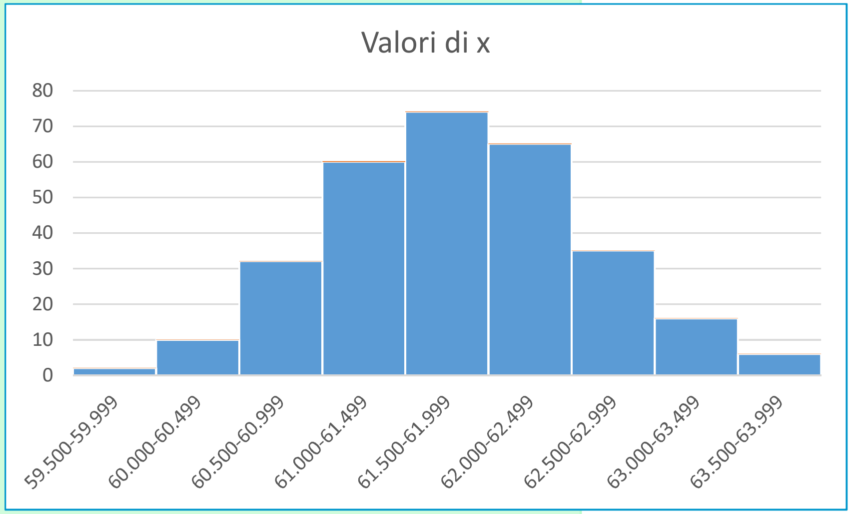
Si estraggono tutti i 300 campioni possibili di numerosità 30 dei manager di Amazon. Per ciascun campione calcolo il reddito medio. In questo modo otteniamo la distribuzione campionaria di $\bar{x}$.

Numero del campione	Media campionaria ($\bar{x}$)
1	61.214
2	62.370
3	61.486
4	61.530
…	……
300	61.822

Esempio

Salario annuale medio	Frequenza	Frequenza relativa
59.500-59.999	2	0,007
60.000-60.499	10	0,033
60.500-60.999	32	0,107
61.000-61.499	60	0,200
61.500-61.999	74	0,247
62.000-62.499	65	0,217
62.500-62.999	35	0,117
63.000-63.499	16	0,053
63.500-63.999	6	0,020
	300	1

L'istogramma della frequenza relativa dei valori di x da 300 campioni casuali semplici ha una forma a campana approssimativamente.



Teorema del limite centrale

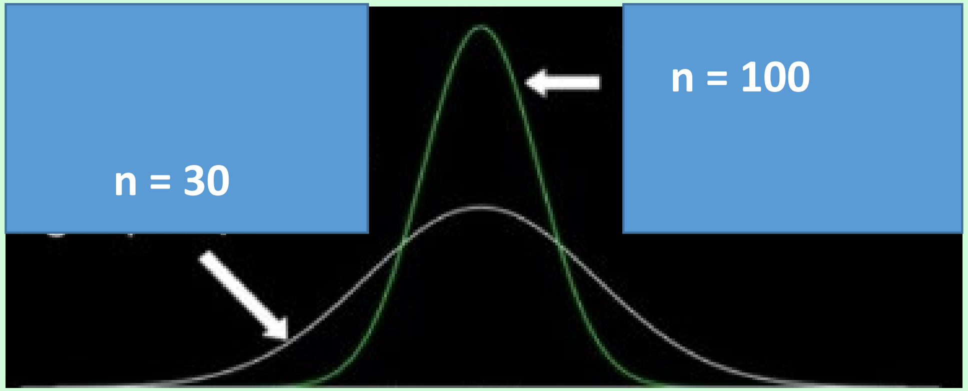
Sia data una popolazione avente media μ e varianza σ^2, e da essa si estraggano campioni casuali di ampiezza n, indicando con $\bar{x}$ la media campionaria, la variabile

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

è una variabile aleatoria la cui distribuzione tende alla distribuzione normale standardizzata per $n \longrightarrow \infty$

Relazione tra la dimensione del campione e la distribuzione campionaria di $\overline{x}$

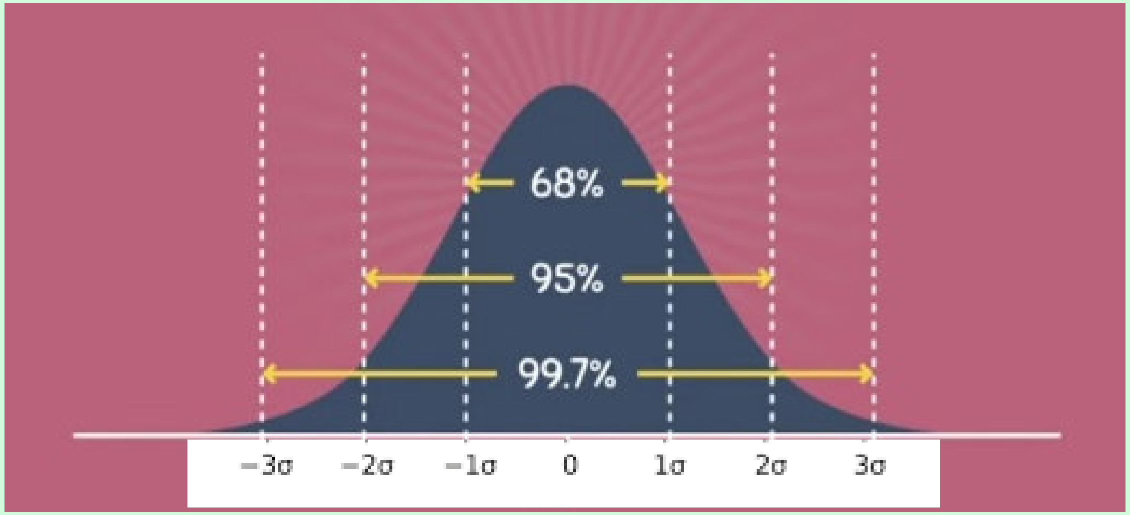
Poiché la distribuzione campionaria con ad es. n = 100 ha un errore standard più piccolo, i valori di $\overline{x}$ hanno una minore variabilità e tendono ad essere più vicini alla media della popolazione rispetto ad $\overline{x}$ con n = 30.



Teorema del limite centrale

Sapendo che la distribuzione della media campionaria è normale possiamo utilizzare la **regola empirica** della distribuzione normale.

Si può calcolare una probabilità associata a un campione.



Esempio

Una popolazione ha una media di 220 e una deviazione standard di 60. Viene selezionato un campione di 100 unità e la media campionaria viene utilizzata per stimare la media della popolazione.

1- La media campionaria di $\overline{x}$ è lo stimatore puntuale della media della popolazione μ?

2- Qual è la deviazione standard di $\overline{x}$?

3- Mostrare la distribuzione campionaria di $\overline{x}$.

4. Cosa mostra la distribuzione campionaria di $\overline{x}$?

Soluzione

1- Si

2- Qual è la deviazione standard di $\bar{x}$?

$= 60/\sqrt{100} = 6$

3- Mostrare la distribuzione campionaria di $\bar{x}$.

Approssimativamente una normale con $\bar{x} = 220$ e $\sigma_{x} = 6$

4. Cosa mostra la distribuzione campionaria di $\bar{x}$?

La distribuzione di probabilità di $\bar{x}$.

Esempio

Assumiamo che la deviazione standard sia $\sigma = 15$.
Calcolare l'errore standard della media per campioni di numerosità 50, 100, 150,200.

Cosa si può dire sulla dimensione dell'errore standard della media all'aumentare della dimensione del campione?

Soluzione

- 2,12
- 1,5
- 1,22
- 1,06

σ_x diminuisce al crescere di n.

Un campione più grande fornisce una maggiore probabilità che la media campionaria sia compresa entro una certa distanza dalla media della popolazione.

Esempio

Una popolazione ha una media di 220 e una deviazione standard di 60. Viene selezionato un campione di 100 unità e la media campionaria viene utilizzata per stimare la media della popolazione.

1- Qual è la probabilità che $\bar{x}$ sia compresa entro ± 6 rispetto alla media della popolazione?
2- Qual è la probabilità che $\bar{x}$ sia compresa entro ± 12 rispetto alla media della popolazione?

Soluzione

Essendo il campione > 30 possiamo utilizzare la tavola delle probabilità normale standardizzata:

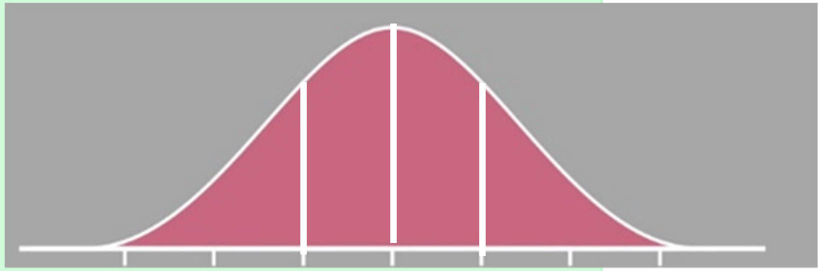
$$z = \frac{x - \mu}{\sigma} \quad \Rightarrow \quad z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$\sigma_x = 60/ \sqrt{100} = 6$

Cerco l'intervallo $(214 \leq \bar{x} \leq 226)$

Per x 214 = (214-220)/ 6 = - 1
Per x 226 = (226-220)/ 6 = + 1

$(-1 \leq P \leq 1)$



214 220 226

Soluzione

P $\leq 1 = 0{,}3413$

P $\geq -1 = 0{,}3413$

$(214 \leq x \leq 226) =$
$0{,}3413 + 0{,}3413 =$
$0{,}6826$

z	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1809	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	[illegible]	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	[illegible]	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4811
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986

Soluzione

P $\leq$ 2 = 0,4772

P $\geq$- 2 = 0,4772

(208 $\leq$ x$\leq$ 232)=
0,4772 + 0,4772=
0,9544

z	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1809	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4811
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986

Distribuzione campionaria di $\overline{x}$

La distribuzione campionaria di x può essere utilizzata per fornire informazioni probabilistiche sulla differenza tra la media campionaria e la media della popolazione.

Esempio in un'azienda lavorano 2.500 dipendenti, il cui stipendio medio è di 21.800€ mensili. Il direttore del personale ritiene che la media campionaria possa essere utilizzata se compresa entro i 400€ dalla media della popolazione.

La domanda in termini probabilistici è: qual è la probabilità che la media campionaria calcolata utilizzando un campione casuale semplice di 30 dipendenti sia compresa entro i 400€ dalla media della popolazione. La deviazione standard della popolazione è 2.000.

Esempio

Si considera un intervallo di probabilità 21.400€ $\leq$ 21.800€ $\leq$ 22.200€

1- Calcolo l'errore standard = 2000/ $\sqrt{30}$= 365,1

z = (21.400-21.800)/ 365,1 = 1,09
z= (22.200-21.800)/365,1 = - 1,09

= 0,3621+0,3621= 0,7242

Un campione casuale semplice di 30 dipendenti ha una probabilità del 72% di fornire una media campionaria che si trovi entro la soglia consentita entro 400€.

z	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1809	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767

Esercizio

I risultati di un'indagine indicano che l'errore standard della media è 20. La deviazione standard della popolazione è 500.

1- Quanto è grande il campione utilizzato per questa indagine?

Soluzione

1- Quanto è grande il campione utilizzato per questa indagine?

$20 = 500/x$

$20x = 500$ ⇨ $500/20 = 25$ ⇨ $n = 625$