Y557 - ESAMI DI MATURITÁ SCIENTIFICA SPERIMENTALE

PIANO NAZIONALE INFORMATICA

Tema di: MATEMATICA

3. Una macchina produce barre di acciaio a sezione circolare la cui lunghezza ottimale dovrebbe essere di 5 metri ed il diametro della sezione di 4 centimetri. Le barre effettivamente prodotte, che si suppongono tra loro indipendenti, hanno una lunghezza aleatoria con distribuzione normale di media $m_1 = 5$ m e scarto standard $\sigma_1 = 4$ cm. Il diametro della sezione è una variabile aleatoria, indipendente dalla precedente, e con distribuzione normale di media $m_2 = 4$ cm e scarto standard $\sigma_2 = 0{,}8$ cm.
Una generica barra prodotta può essere direttamente venduta senza modifiche se la sua lunghezza è compresa tra 4,95 m e 5,05 m e la sua sezione tra 2,8 cm e 5,2 cm.
La tavola della funzione di ripartizione della distribuzione normale standardizzata è, per alcuni valori, la seguente:

Ascissa: x	**F(x)**
- 1,50	0,067
- 1,45	0,074
- 1,35	0,089
- 1,25	0,106
- 1,15	0,125
- 1,05	0,147
- 0,95	0,171
+ 0,95	0,829
+ 1,05	0,853
+ 1,15	0,875
+ 1,25	0,894
+ 1,35	0,912
+ 1,45	0,927
+ 1,50	0,933

Il candidato:

a) verifichi che la probabilità p di poter mettere in vendita senza modifiche una generica barra prodotta è 0,68;
b) indicata con f_n la frequenza relativa alle barre direttamente vendibili su n barre prodotte, esprima, in funzione di p, la numerosità n necessaria perché la probabilità che f_n disti da p più di 0,05 sia non superiore a 0,05;
c) dato il valore di p rilevato in a), se su 2000 barre prodotte 1000 risultano non direttamente vendibili, dica se si può sospettare che la macchina non funzioni più secondo lo standard riportato sopra, se, cioè, il risultato ottenuto risulta a priori poco probabile (probabilità inferiore a 0,05) subordinatamente alle modalità di funzionamento della macchina, come indicato;
d) descriva una procedura che consenta di calcolare la probabilità di ottenere la prima barra direttamente vendibile solo all'n-esima prova, al variare di p e di n, e la codifichi in un linguaggio di programmazione conosciuto.

Durata massima della prova: 5 ore.
E' consentito l'uso della calcolatrice tascabile non programmabile.
Non è consentito lasciare l'Istituto prima che siano trascorse 3 ore dalla dettatura del tema.

Soluzione.

La funzione di distribuzione normale è $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, dove
μ è $la\ media\ \ e\ \sigma\ la\ deviazione\ standard$; il suo Il grafico è la curva di Gauss.
Se X è un valore della variabile aleatoria, la probabilità che X appartenga all'intervallo $[a; b]$ è

$$P[a \le X \le b] = \int_a^b f(x)dx$$

$Geometricamente$ è $l'area\ sottesa\ dalla\ curva\ sull'intervallo [a; b]$
Inoltre,

$$P[X \le b] = \int_{-\infty}^{b} f(x)dx \quad P\ [X \ge a] = \int_a^{+\infty} f(x)dx$$

In particolare

$$\int_{-\infty}^{\mu} f(x)dx\ = \int_{\mu}^{+\infty} f(x)dx = 0{,}5 \qquad \int_{-\infty}^{+\infty} f(x)dx = 1$$

La tabella allegata si riferisce alla distribuzione normale standardizzata (di media uguale a 0 e deviazione standard uguale 1), $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$, risultato della trasformazione

$$\begin{cases} z = x' = \dfrac{x-\mu}{\sigma} \\ y' = \sigma y \end{cases}$$

(la trasformazione è una particolare affinità che conserva le aree)

La tabella fornisce i valori dell'area sottesa dalla curva sull'intervallo $]-\infty; x]$

$$F(x) = \int_{-\infty}^{x} f(z)dz \quad \text{(funzione di ripartizione)}$$

m$entre\ l'area\ sottesa\ dalla\ curva\ sull'intervallo [a; b]\ sar$è $uguale\ a\ \ F(b) - F(a)$

La variabile aleatoria $X_l$che descrive la lunghezza di una barra ha densità di probabilità

$$f_1(x) = \frac{1}{\sqrt{2\pi} \cdot 4} e^{-\frac{1}{2}\left(\frac{x-500}{4}\right)^2}$$

mentre la variabile aleatoria X_d che descrive il diametro della sezione di una barra ha densità di probabilità

$$f_2(x) = \frac{1}{\sqrt{2\pi} \cdot 0.8} e^{-\frac{1}{2}\left(\frac{x-4}{0.8}\right)^2}$$

a) *La probabilità che una barra prodotta abbia una lunghezza compresa tra 495 cm e 505 cm è uguale a*

$$\int_{495}^{505} f_1(x)dx$$

Passando ai corrispondenti valori standardizzati , si osserva che 495 corrisponde a $\frac{495-500}{4} = -1{,}25$ mentre 505 corrisponde a $\frac{505-500}{4} = 1{,}25$

$$P[-1{,}25 \leq X_l \leq 1{,}25] = F(1{,}25) - F(-1{,}25) \approx 0{,}788$$

La probabilità che una barra prodotta abbia una sezione di diametro compreso tra 2,8 cm e 5,2 cm è uguale a

$$\int_{2,8}^{5,2} f_2(x)dx$$

Passando ai corrispondenti valori standardizzati , si osserva che 2,8 corrisponde a $\frac{2,8-4}{0,8} = -1{,}5$ mentre 5,2 corrisponde a $\frac{5,2-4}{0.8} = 1{,}5$

$$P[-1{,}25 \leq X_d \leq 1{,}5] = F(1{,}5) - F(-1{,}5) \approx 0{,}866$$

La probabilità p di poter mettere in vendita senza modifiche una generica barra prodotta risulta dal prodotto delle probabilità di due eventi indipendenti. Pertanto

$$\boldsymbol{p = 0,788 \cdot 0,866 \approx 0.68}$$

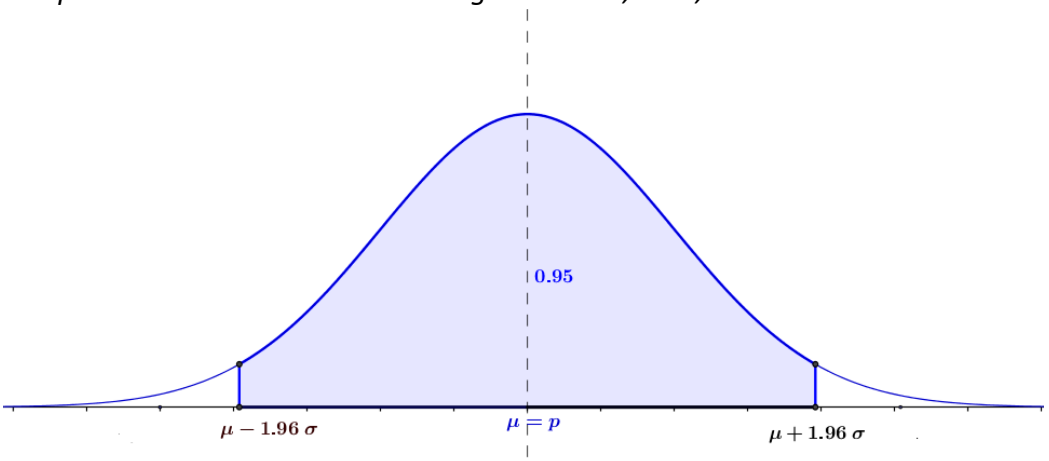
b) *La variabile aleatoria X="numero di barre direttamente vendibili tra n barre prodotte " è binomiale, con media uguale a np e varianza $np(1-p)$*
Analogamente la frequenza relativa $f_n = \frac{X}{n}$ è una variabile binomiale, con media uguale a p e varianza
$\frac{n \cdot p(1-p)}{n^2} \rightarrow \sigma = \sqrt{\frac{p(1-p)}{n}}$.
Si chiede per quali valori di n risulti $P(\left|\frac{X}{n} - p\right| > 0{,}05) \leq 0{,}05$

Primo metodo

Approssimiamo la binomiale con la normale, per valori sufficientemente grandi di n.
E' naturale aspettarsi che i valori di una variabile aleatoria X si dispongano intorno alla media μ e che una unità di misura appropriata per valutare il discostamento dalla media sia la deviazione standard σ
Ricordando che, per una variabile aleatoria a distribuzione normale risulta

$$95{,}0\% = P\{\mu - 1{,}96\,\sigma < X < \mu + 1{,}96\,\sigma\}$$

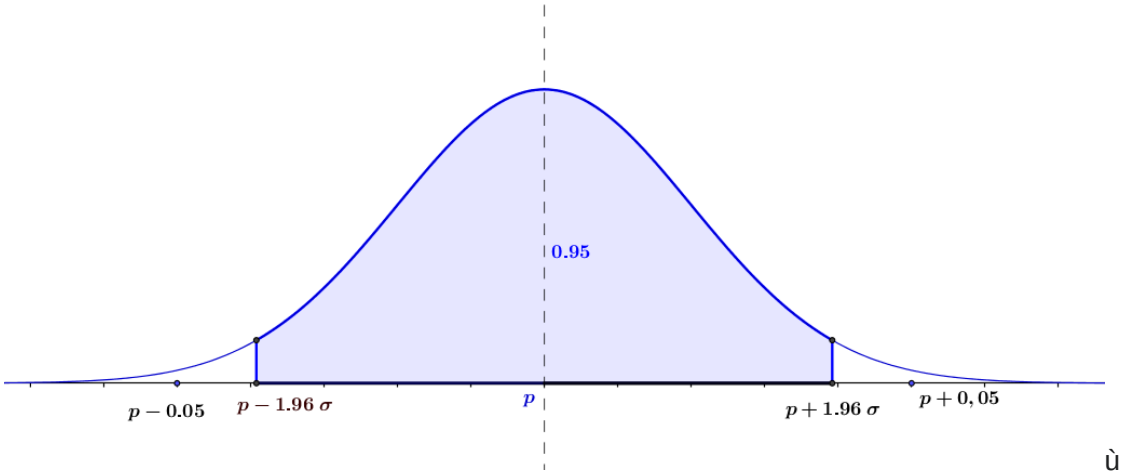
Questo significa che il valore X è interno all'intervallo $\{\mu - 1{,}96\,\sigma\ ;\ \mu + 1{,}96\,\sigma\}$ *con probabilità 95% ed esterno allo stesso intervallo con probabilità 5%.*
Geometricamente : l'area sottesa dalla curva nell'intervallo considerato è uguale a 0,95; l'area complessiva delle due <<code >> è uguale a 1-0,95=0,05



Osserviamo che $P\left(\left|\frac{X}{n} - p\right| > k\right)$ corrisponde alla probabilità che il valore $\frac{X}{n}$ cada all'esterno dell'intervallo $[p - k; p + k]$ mentre $P\left(\left|\frac{X}{n} - p\right| \leq k\right)$ corrisponde alla probabilità che il valore $\frac{X}{n}$ cada nello stesso intervallo (evento complementare)

Vogliamo che $P\left(\left|\frac{X}{n} - p\right| > 0{,}05\right) \leq 0{,}05$ cioè

il valore $\frac{X}{n}$ ***deve essere esterno all'intervallo*** $[p - 0{,}05; p + 0{,}05]$ ***con probabilità non superiore al 5% .***
I valori $p - 0{,}05$ ***e*** $p + 0{,}05$ ***devono cadere, quindi, nelle due code della gaussiana, con valori critici*** $p - 1{,}96\,\sigma$ ***e*** $p + 1{,}96\sigma$.



ù

Pertanto, si impone che

$$p - 1{,}96\sqrt{\frac{p(1-p)}{n}} \geq p - 0{,}05 \qquad p + 1{,}96\sqrt{\frac{p(1-p)}{n}} \leq p + 0{,}05$$

da cui

$$\boldsymbol{n \geq \left(\frac{1{,}96}{0{,}05}\right)^2 (p - p^2)}$$

Secondo metodo

IL teorema di Bienaymé-Cebycev afferma che la probabilità che un dato x si discosti dalla media non meno di uno scarto pari a *k* volte σ è al massimo $\frac{1}{k^2}$

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Nel caso in questione dobbiamo stimare la probabilità che la variabile aleatoria $\frac{X}{n}$ si discosti dalla media p per più di un valore fissato $t = 0{,}05$.

Essendo quindi $k^2 = \left(\frac{t}{\sigma}\right)^2 = \frac{n(0{,}05)^2}{p(1-p)}$, $affinché\ risulti\ P\left(\left|\frac{X}{n} - p\right| > 0{,}05\right) \leq 0{,}05$, dobbiamo imporre che

$$\frac{p(1-p)}{n(0{,}05)^2} \leq 0{,}05 \rightarrow \boldsymbol{n \geq \frac{(p - p^2)}{(0{,}05)^3}}$$

Osservazione

Poiché, come si verifica facilmente, $\left(\frac{1{,}96}{0{,}05}\right)^2 < \frac{1}{(0{,}05)^3}$, il risultato ottenuto con la disuguaglianza di Cebycev è meno preciso e spesso fornisce una stima grossolana della probabilità richiesta. E' utile quando non si hanno informazioni sulla natura della distribuzione ma si conoscono la media e la varianza.

Sostituendo a p , ad esempio, il valore determinato nel punto a, si trova

$$n \geq \left(\frac{1{,}96}{0{,}05}\right)^2 (p - p^2) \rightarrow n \geq 334{,}37\ nel\ primo\ caso$$

$$n \geq \frac{(p - p^2)}{(0{,}05)^3} \rightarrow \quad n \geq 1740{,}8 \quad nel\ secondo\ caso$$

Stimiamo che un valore di frequenza relativa $\frac{X}{n}$ ***abbia dalla media uno scarto (in modulo) superiore a 0,05 con probabilità non superiore al 5% , per*** $n \geq \left(\frac{1,96}{0,05}\right)^2 (p - p^2)$ ***se accettiamo l'approssimazione della binomiale con la normale, altrimenti ci limitiamo al secondo risultato ottenuto con la*** **disuguaglianza di Cebycev**

c) Si osserva che,su 2000 barre prodotte, il numero di quelle invendibili sono 1000, numero inferiore a quello previsto con una media del 68%.

In continuità col punto b possiamo affermare che è stato rilevato il valore $f_{2000} = 0,5$ il cui scarto, in modulo, dalla media $p= 0,68$ è pari a $|0,8 - 0,68| = 0,18 > 0,05$.

Si deve stabilire se lo scostamento del risultato empirico da quello atteso è dovuto al "caso", o al fatto la macchina non funzioni più secondo lo standard iniziale.

Primo metodo

Il criterio, che permette di dare una risposta con un grado di fiducia del 5%, è suggerito proprio dal risultato del punto b .

1. *Il valore di frequenza* $f_n = \frac{X}{n} = 0,5$ *ha uno scarto rispetto alla media* $p= 0,68$ *pari a* $|0,8 - 0,68| = 0,18 > 0,05$, *quindi rientra nell'analisi del punto b*

2. *Il suddetto valore ha una probabilità minore del 5% se* $n \geq \left(\frac{1,96}{0,05}\right)^2 (p - p^2)$

3. *Sostituendo* $n=2000$, $p=0,68$ *si trova*

$2000 \geq 334,37$ ***Verificata***

Il valore di frequenza $f_n = \frac{X}{n} = 0,5$ differisce in modo significativo dal valore atteso, pertanto , con un grado di fiducia del 5%, possiamo affermare che la macchina non funziona più secondo lo standard riportato nella parte iniziale del testo

Secondo metodo

Si può anche rispondere al quesito senza tener conto del risultato ottenuto nel punto b, seguendo la procedura classica dei problemi di verifica delle ipotesi.

Risultato del test statistico : su 2000 barre prodotte il numero di quelle invendibili sono 1000→ $\frac{X}{n} = f = 0,5$

Sia H_0 $l'ipotesi\ nulla$:

$$Non\ c'\text{è}\ differenza\ significativa\ tra\ la\ frequenza\ osservata\ e\ quella\ attesa;$$

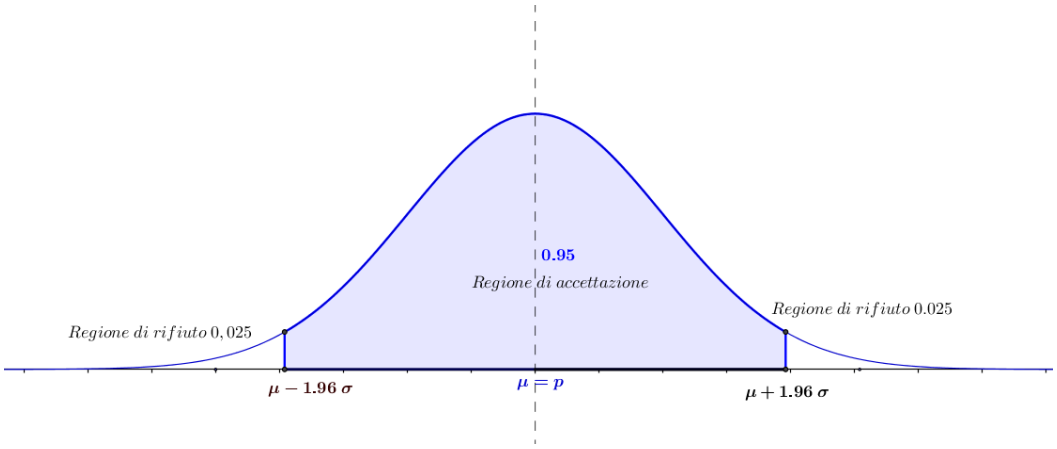
$$la\ macchina\ funziona\ normalmente$$

Sia H_1 $l'ipotesi\ alternativa$:

$$C'\text{è}\ differenza\ \ significativa\ tra\ la\ frequenza\ osservata\ e\ quella\ attesa;$$

$$la\ macchina\ non\ funziona\ pi\text{ù}\ secondo\ lo\ standard\ iniziale$$

La Regione di rifiuto**,** con livello di significatività 5% è quella indicata nella figura seguente, dove è stato preso in esame un test bilaterale.
Si accetta, cioè che, nell'ipotesi alternativa , la macchina produca o un numero minore o un numero maggiore di prodotto vendibili, in accordo col fatto che il testo non fornisce alcuna precisazione in proposito, anche se il buon senso suggerisce che ciò che interessa maggiormente è un eventuale malfunzionamento.



$$Se\ passiamo\ ai\ valori\ standardizzati\ \ i\ valori\ critici\ saranno\ -1{,}96\ e\ \ +1{,}96$$

Valore standardizzato $z = \frac{0{,}5-0{,}68}{\sqrt{\frac{0{,}68\cdot 0{,}32}{2000}}} \approx -17{,}3 \ll -1.96$

Il risultato del test $|z| > 1{,}96$ fa rifiutare l'ipotesi nulla con un livello di significatività de 5%

d) la probabilità di ottenere la prima barra direttamente vendibile solo all'n-esima prova (primo successo all'n-sima prova) si ottiene come prodotto delle probabilità di n eventi indipendenti;

per n-1 volte si ha un insuccesso $(probabilità = q = 1 - p)$. e all'n-sima si ha un successo $(probabilità = p)$.

La distribuzione è geometrica

1	**2**	**3**	**4**	**5**	.	.	.	**n**
p	pq	pq^2	pq^3	pq^4		.	.	pq^{n-1}

Si ottengono n termini di una successione definita ricorsivamente nel modo seguente

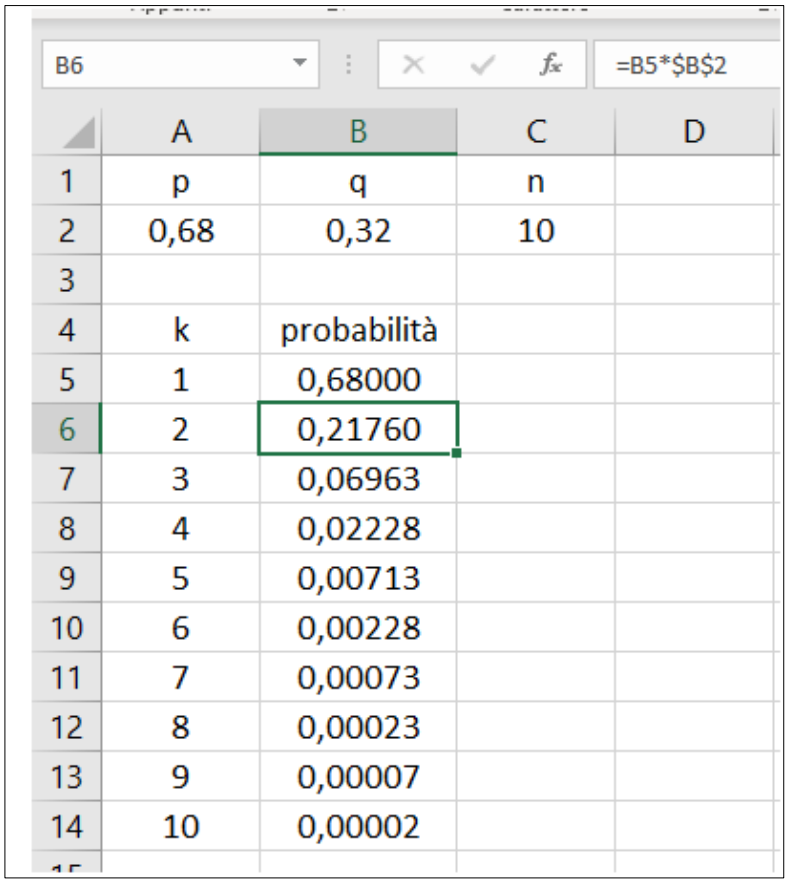
$$\begin{cases} P_1 = & p & n = 1 \\ P_n = & q \cdot P_{n-1} & n > 1 \end{cases}$$

Esempi di utilizzo dell'algoritmo

Linguaggio PASCAL

```
program geometrica;
var  p:real; n:integer
      function successione (x:real; y:integer):real;
begin
   if n=1   then
     successione:=x
   else
successione:=(1-x)* successione (x,y-1);
end;
begin
writeln("introduci il valore della probabilità maggiore di 0 e minore di 1"); readln(p);
writeln("introduci il numero delle prove); readln(n);
writeln("la probabilità di avere il primo successo  alla n-sima prova è");
writeln(successione (p,n):8:5);
end.
```

EXCEL



	A	B	C	D
1	p	q	n	
2	0,68	0,32	10	
3				
4	k	probabilità		
5	1	0,68000		
6	2	0,21760		
7	3	0,06963		
8	4	0,02228		
9	5	0,00713		
10	6	0,00228		
11	7	0,00073		
12	8	0,00023		
13	9	0,00007		
14	10	0,00002		