Metodi Matematici e Statistici, modulo di statistica, Ing. Gestionale, prova scritta del 16/1/04

1) Un'azienda farmaceutica esamina l'efficacia di un proprio farmaco in commercio da un po' di tempo, per studiare poi eventuali modifiche. Tale farmaco promette di portare un certo indicatore (misurabile con analisi del sangue) al valor medio 420, con varianza 40. Supponiamo di credere nella varianza, ma di avere dubbi circa la media dichiarata. L'azienda raccoglie allora dei dati da alcuni centri ospedalieri e rileva quanto segue: relativamente ad 80 persone curate con quel farmaco, si è trovato il valor medio $\bar{x} = 430$ per quell'indicatore. Al 95%, si può ritenere corretto il valor medio 420 dichiarato?

2) Supponiamo che 430 sia un valore pericoloso (per la salute), quindi interessi valutare la capacità del test precedente di accorgersi che la media vera è 430 invece che 420. Che potenza ha il test precedente da questo punto di vista?

3) L'azienda mette a punto delle modifiche al farmaco e vuole stimare la nuova media al 99% con precisione pari a 5. Quanti esperimenti deve effettuare?

4) Si osserva in un'ospedale che tale farmaco ha probabilità p di indurre alcune complicanze di carattere digestivo. Nel tentativo di verificare questa osservazione, vengono esaminate più attentamente N persone in un'altro ospedale. Se fosse vero che nascono tali complicanze con tale probabilità, che probabilità ci sarebbe di osservare un numero $\leq k$ di complicanze negli N pazienti esaminati? Rispondere nei seguenti casi: $p = 0.2$, $N = 5$, $k = 2$, oppure $p = 0.05$, $N = 50$, $k = 4$, oppure infine $p = 0.3$, $N = 100$, $k = 35$. Ci sono vari modi (alcuni esatti, altri approssimati) di rispondere; scegliere in ciascun caso quello che sembra più adatto.

5) Se in una generica città il nuovo farmaco vende bene, l'azienda non fa particolari operazioni. Se invece non vende bene, l'azienda si chiede se intensificare la pubblicità fatta dai suoi rappresentanti, e decide di farlo nel 40% dei casi. Supponiamo che questa maggior attività pubblicitaria costi all'azienda, in una generica città, 10000 euro in più. Se si osserva che il farmaco vende bene solo nel 20% delle città, qual'è il costo medio in più che l'azienda deve sostenere?

Soluzioni.

1) Si esegue un test per la media, usando quantili gaussiani. L'ipotesi nulla è $\mu_0 = 420$. Si calcola $z = \sqrt{n}\frac{\bar{x}-\mu_0}{\sigma} = \sqrt{80}\frac{10}{\sqrt{40}} = 14.142$. Esso supera abbondantemente il quantile $q_{1-\frac{\alpha}{2}}$ con $1-\alpha = 95\%$, ovvero $q_{1-\frac{\alpha}{2}} = q_{0.975} = 1.96$. Quindi rigettiamo l'ipotesi.

2) La potenza è $1-\beta$ dove

$$\begin{aligned}\beta &= \Phi\left(\sqrt{n}\frac{\mu_0-\mu}{\sigma} + q_{1-\frac{\alpha}{2}}\right) - \Phi\left(\sqrt{n}\frac{\mu_0-\mu}{\sigma} - q_{1-\frac{\alpha}{2}}\right)\\ &= \Phi\left(-14.142 + 1.96\right) - \Phi\left(-14.142 - 1.96\right) \sim 0.\end{aligned}$$

Quindi la potenza è circa 1.

3) Ora $1-\alpha = 99\%$, ovvero $q_{1-\frac{\alpha}{2}} = q_{0.995} = 2.58$. L'errore δ è pari a $\frac{\sigma q_{1-\frac{\alpha}{2}}}{\sqrt{n}} = \frac{\sqrt{40}\cdot 2.58}{\sqrt{n}}$ quindi deve essere

$$\frac{16.317}{\sqrt{n}} \leq 5$$

ovvero

$$\sqrt{n} \geq \frac{16.317}{5} = 3.2634$$

quindi $n = 11$ è sufficiente.

4) In generale, è corretto usare una binomiale $B(N,p)$ per descrivere il numero di complicanze, numero aleatorio che indicheremo con X. Infatti, per ogni paziente esaminato, introduciamo una v.a. di Bernoulli di parametro p che vale uno se il paziente manifesta una complicanza. Pertanto $X = X_1 + ... + X_N$, che assumendo l'indipendenza delle X_i è una $B(N,p)$.

A volte però può convenire l'uso di approssimazioni.

Nel primo caso $p = 0.2$, $N = 5$, $k = 2$ usiamo la binomiale $B(5, 0.2)$:

$$\begin{aligned}P(X \leq 2) &= \sum_{j=0}^{2}\binom{N}{j}p^j(1-p)^{N-j}\\ &= 0.8^5 + 5\cdot 0.2\cdot 0.8^4 + \frac{5\cdot 4}{2}0.2^2\cdot 0.8^3 = 0.942.\end{aligned}$$

Nel seondo caso $p = 0.05$, $N = 50$, $k = 4$, abbiamo a he fare con eventi rari ($p = 0.05$), in discreta quantità ($N = 50$), quindi è ragionevole applicare l'approssimazione di Poisson: $X \sim P(2.5)$. Vale allora

$$P(X \leq 4) \sim \sum_{j=0}^{4} e^{-2.5} \frac{2.5^j}{j!}$$
$$= e^{-2.5} \left(1 + 2.5 + \frac{2.5^2}{2} + \frac{2.5^3}{6} + \frac{2.5^4}{24} \right) = 0.891.$$

Il calcolo con la Poisson è in effetti un po' più agevole che con la binomiale.

Infine, net terzo caso $p = 0.3$, $N = 100$, $k = 35$, conviene usare l'approssimazione normale, essendo k, ed anche N, alto:

$$P(X_1 + ... + X_N \leq 35) = P\left(\frac{X_1 + ... + X_N - Np}{\sqrt{N}\sqrt{p(1-p)}} \leq \frac{35 - Np}{\sqrt{N}\sqrt{p(1-p)}} \right)$$
$$\sim \Phi(1.09) = 0.862$$

essendo $\frac{35-Np}{\sqrt{N}\sqrt{p(1-p)}} = 1.09$.

5) In ogni città ci sono solo due possibilità: o vengono spesi i 10000 euro oppure no. La probabilità che vengano spesi è, per la formula di fattorizzazione,

$$0.4 \cdot 0.8 + 0 \cdot 0.2 = 0.32.$$

Quindi la spesa media è di

$$0.32 \cdot 10000 + 0.68 \cdot 0 = 3200$$

euro.