1 Esercizi di Metodi Statistici, Ing. Gestionale

Esercizio 1. Calcolare la probabilità di effettuare un ambo giocando due numeri su una ruota (90 numeri, 5 estratti diversi, l'ordinamento non conta).

Esercizio 2. Calcolare la probabilità di trovare k teste in n lanci.

Esercizio 3. Calcolare la probabilità di totalizzare k teste al momento dell'n-esimo lancio.

Esercizio 4. Calcolare la probabilità di ricevere 5 visite di conoscenti nello stesso giorno, se si conoscono 20 persone (indipendenti), ciascuna che ci fa visita giornalmente con probabilità 0.2.

Esercizio 5. Un produttore produce pezzi funzionanti al 95%. Un sistema di controllo li esamina e ferma il 90% di quelli non funzionanti, mentre lascia passare tutti quelli funzionanti. Qual'è la probabilità che il sistema lasci passare un pezzo?

Esercizio 6. Consideriamo un sistema composto da due componenti c_1 e c_2 in parallelo. Supponiamo che c_1 funzioni con probabilità p_1 e c_2 funzioni con probabilità p_2. Calcolare la probabilità che il sistema non funzioni. Supponiamo tacitamente in questo e negli esercizi successivi che l'eventuale mal funzionamento di una componente sia indipendente dall'altra.

Esercizio 7. Due strade portano a destinazione, ma ciascuna può essere interrotta con probabilità 0.01. Calcolare la probabilità di poter raggiungere la destinazione.

Esercizio 8. In un sistema composto da due componenti in serie, ciascuna può essere rotta con probabilità 0.1. Calcolare la probabilità che il sistema funzioni.

Esercizio 9. Se un termostato in un impianto funziona correttamente con probabilità 0.95, qual'è la probabilità che un sistema composto da 2 termostati in serie non funzioni correttamente?

Esercizio 10. Si pensi di svolgere un controllo della qualità dei pezzi prodotti da una certa industria. Supponiamo di sapere in partenza che il generico pezzo ha una probabilità 0.01 di essere rotto. Calcolare la probabilità di trovare 5 pezzi rotti su 1000 esaminati. Fornire anche una formula approssimata.

Esercizio 11. Uno strumento elettronico C contiene varie componenti di cui una fondamentale A. Se A funziona sotto un certo limite di tolleranza, C ha probabilità 0.1 di funzionare male, altrimenti ha probabilità 0.05 di funzionare male. D'altra parte, la probabilità che A funzioni sotto il limite di tolleranza 0.01. Calcolare la probabilità che C funzioni male. Osservato un mal funzionamento, calcolare la probabilità che esso sia dovuto ad A.

Esercizio 12. Sottoponiamo il 30% della produzione ad un nuovo trattamento N. Da osservazioni sperimentali risulta che il 5% dei prodotti non trattati con N è scadente, mentre tra quelli trattati lo è il 3%. Preso un pezzo ed osservato che è scadente, che probabilità c'è che abbia ricevuto il trattamento?

Esercizio 13. Sia $f(x)$ la densità uniforme su $[0, a]$. Dare la formula per $f(x)$ e calcolarne la funzione di ripartizione $F(x)$.

Esercizio 14. Date $X_1, ..., X_n$ Bernoulli indipendenti di parametro p, posto $S_n = X_1 + ... + X_n$, Calcolare $P(S_n = k)$.

Esercizio 15. Calcolare media e varianza di una Bernoulli di parametro p, e di una binomiale $B(n, p)$.

Esercizio 16. Mostrare che se A e B sono eventi indipendenti, allora lo sono anche A e $\overline{B}$.

Esercizio 17. Certi oggetti vengono sottoposti o meno ad un trattamento. Di 8 trattati, 5 risultano di buona qualità, 3 no. Di 10 non trattati, 2 risultano di buona qualità, 8 no. trovare una valutazione quantitativa della convinzione intuitiva che il trattamento abbia effetti positivi (in particolare, calcolare la probabilità di avere simili risultati in assenza completa di trattamento).

Esercizio 18. Si ritiene che mediamente 10 italiani su 100 comprino un certo prodotto. Dopo un certo periodo di vendite si scopre che su 500 clienti solo 35 hanno scelto quel prodotto. Ritenete verosimile la previsione iniziale del 10%? Si richiede solamente di impostare il conto che risponderebbe a questo quesito, senza svolgere i calcoli troppo lunghi (salvo si usi un'approssimazione opportuna...).

Esercizio 19. Calcolare media, varianza e funzione generatrice di una v.a. di Poisson di parametro λ.

Esercizio 20. Se X_n è $B(n,p)$ e X è $P(\lambda)$, con $\lambda = np$, mostrare che $\Phi_{X_n} \to \Phi_X$ per $n \to \infty$.

Esercizio 21. Se X, Y sono $P(\lambda)$ e $P(\mu)$ indipendenti, mostrare che $X + Y$ è $P(\lambda + \mu)$.

Esercizio 20. Due ditte producono un certo tipo di macchine. La ditta A ne ha prodotte 6 e venduta 1, la ditta B ne ha vendute 4 su 9. Si può affermare che la ditta B ha maggior probabilità di vendita? Fornite un numero che dia un'indicazione quantitativa della vostra opinione.

Esercizio 21. Un venditore ritiene che un certo modello di automobile abbia successo se viene acquistato dal 30% dei clienti. Dopo alcuni mesi registra che 2 su 10 hanno acquistato quel modello. Può concludere che il modello non ha successo? Fornite un numero che dia un'indicazione quantitativa della vostra opinione.

Esercizio 22. Un dispositivo è basato su 3 relais in parallelo. Ogni relais può presentare 2 tipi di malfunzionamento: ossidazione, bloccaggio. L'ossidazione non fa agire il relais, il bloccaggio lo fa rimanere azionato. Calcolare la probabilità che il sistema funzioni. Si assuma che la probabilità di ossidazione è 0.007, quella di bloccaggio è 0.003.

Esercizio 23. Sia X una v.a. esponenziale di parametro λ. Calcolare funzione generatrice, media e varianza.

Esercizio 24. Un sistema di comunicazione funziona con due server in parallelo, A e B. Ciascuno ha probabilità 0.999 di essere attivo. L'input può essere schematizzato in due fasce: la fascia M d'intensità massima (ora di punta) e quella N di intensità normale. Quando l'input è N ed entrambi i server sono attivi, la probabilità di trasmissione senza perdite d'informazione è 0.9999, mentre sotto input M è solo 0.980. Quando però uno solo dei due server è attivo, esse si riducono rispettivamente a 0.60 e 0.35. Ovviamente, quando poi entrambi i server sono disattivi, essa si riduce a zero. Stimando che sia 0.15 la frequenza di input massima e 0.85 quella normale, calcolare la probabilità che, in un generico istante scelto a caso, il sistema trasmetta senza perdite d'informazione.

Esercizio 25. Date $X_1, ..., X_n$ indipendenti con uguali medie μ e varianze σ^2, posto $S_n = X_1 + ... + X_n$, calcolare media e varianza di S_n, osservando dove serve l'ipotesi d'indipendenza.

Esercizio 26. Nelle ipotesi precedenti, trovare media e varianza di $Z_n = \frac{X_1+...+X_n-n\mu}{\sigma\sqrt{n}}$. Stabilire inoltre che variabile è Z_n se si assume che le X_i siano $N(\mu, \sigma^2)$, mentre dire cosa si sa di Z_n in generale.

Esercizio 27. Calcolare la probabilità di avere ≤ 45 successi su 500 prove indipendenti, sapendo che la probabilità di successo in una prova è del 10%.

Esercizio 28. Sia R l'affidabilità (=probabilità di buon funzionamento) di un componente. Stabilire l'affidabilità di un sistema composto da 3 componenti in parallelo, che funziona se almeno due componenti sono funzionanti.

Esercizio 29. Un compratore impone la seguente regola: controlla 100 esemplari a caso di quelli forniti dal venditore, se ne trova almeno 1 scadente, rifiuta la fornitura. Se il venditore decide che può correre un rischio 0.05 di vedersi rifiutare la fornitura, a quale livello di qualità deve produrre? Con "livello di qualità" qui intendiamo la probabilità che un esemplare della sua produzione sia scadente.

Esercizio 30. Data X di tipo $N(0, 1)$, trovare la densità di probabilità della v.a. $Y = |X|$.

Esercizio 31. Data X di tipo $N(0, 1)$, trovare la densità di probabilità della v.a. $Y = X^2$.

Esercizio 32. Svolgere gli esercizi 1.9, 1.10, 2.1, 2.3, 3.4, 3.17, 4.13, 4.16 del libro di testo.

Esercizio 33. Trovare il modello gaussiano corrispondente al campione sperimentale 5.2, 4.8, 5.5, 5.1.

Esercizio 34. Trovare il modello binomiale di numerosità 10 corrispondente al campione 1, 0, 2, 1, 1, 3, 0.

Esercizio 35. Trovare il modello di Poisson corrispondente al campione precedente.

Esercizio 36. Trovare il modello esponenziale corrispondente al campione 3.6, 5.2, 4.8, 5.6.

Esercizio 37. Sia X una gaussiana di media 4 e da essa sia stato estratto il campione 21, 25, 19, 21.5, 20, 24.5. Stimare la media e trovare l'intervallo di confidenza al 95%.

Esercizio 38. Relativamente all'esercizio precedente, trovare la numerosità campionaria che darebbe una precisione (uguale o superiore a) 0.5.

Esercizio 39. Trovare l'intervallo di confidenza dell'esercizio precedente (n.37) nell'ipotesi più realistica di non conoscere la deviazione standard.

Esercizio 40. Il prezzo di un titolo sia (approssimativamente) gaussiano, con deviazione standard 15. sulla base dei suoi valori osservati in 8 giorni diversi, ovvero 410, 431, 415, 445, 407, 435, 421, 403, stimare la media al 98%. Trovare inoltre la numerosità che garantirebbe una precisione pari a 5.

Esercizio 41. Trovare l'intervallo di confidenza dell'esercizio precedente nell'ipotesi più realistica di non conoscere la deviazione standard.

Esercizio 42. Dalla pratica di diversi anni si sa che il valore medio di una caratteristica X di un prodotto è 500. In seguito ad una nuova scoperta si prova ad utilizzare un nuovo metodo di produzione, atto a far aumentare mediamente il valore della caratteristica X. Esaminando 10 elementi della nuova produzione si trovano dei valori $x_1, ..., x_{10}$ aventi media $\overline{x} = 508$ e deviazione $S = 5$. Con significatività 95%, si può affermare che il nuovo metodo di produzione migliora il prodotto?

Esercizio 43. Relativamente all'esercizio precedente, si supponga nota (per semplicità) la varianza $\sigma^2 = 16$ delle grandezze aleatorie studiate. Supponiamo che ci interessi capire se il valore medio della caratteristica studiata è cresciuto di almeno 5 unità, ovvero se ha una nuova media maggiore o uguale a 505. Calcolare la potenza che ha il test appena eseguito di accorgersi di una tale variazione della media.

Esercizio 44. Il tempo di produzione medio previsto per la fabbricazione di un certo oggetto è di 20 giorni. Una nuova metodologia sembra ridurre tale tempo; viene provata su 15 oggetti avendo come risultati certi tempi di produzione $x_1, ..., x_{15}$ aventi media $\overline{x} = 18,5$ giorni e deviazione $S = 2$ giorni. Con significatività 95%, si può affermare che il nuovo metodo rende più veloce la produzione? Indicare con chiarezza le ipotesi H_0 ed H_1 e la regione di rigetto.

Esercizio 45. Supponiamo di sapere che una gaussiana X ha varianza 9. Viene inoltre ipotizzato che abbia media 100. Un campione sperimentale di numerosità 7 ha dato media $\overline{x} = 115$. Al 95% questo contraddice l'ipotesi? Calcolare anche la potenza di questo test di accorgersi di una nuova media pari a 110.

Esercizio 46. Supponiamo di studiare due grandezze aleatorie X ed Y che pensiamo possano essere collegate. Quattro prove sperimentali hanno dato le seguenti coppie di valori (x_i, y_i): $(1, 2)$, $(2, 3)$, $(3, 5)$, $(4, 8)$. Valutare la plausibilità di un legame lineare del tipo $Y = aX + b + \varepsilon$. Stimare a, b, e la varianza di ε. Quanto ci aspettiamo valga Y quando X vale 5, e quale errore in questa valutazione ci aspettiamo al 95%? Corredare la risoluzione di una rappresentazione grafica.

Esercizio 47. Relativamente ai dati dell'esercizio precedente, ritenete più plausibile un legame del tipo $Y = aX^2 + b + \varepsilon$? Rispondere in questo caso a tutte le altre domande dell'esercizio precedente. Trovate che la scelta tra i due modelli abbia rilevanza ai fini della predizione del valore di Y in corrispondenza di $X = 5$?

Esercizio 48. Svolgere gli esercizi 6.1, 6.3, 6.8, 7.1, 7.2, 7.3, 7.4, 7.5, 7.6, 7.7 del libro di testo.

Esercizio 49. Indichiamo con X l'incasso giornaliero di un supermercato. In condizioni stazionarie, supponiamo sia una v.a. gaussiana di media μ. I gestori individuano i seguenti fattori come potenziali correttivi al fine di migliorare gli incassi: la pubblicità P, gli sconti S, la qualità Q della merce. Indicare la struttura generale di un modello di questo problema, adatta ad una successiva progettazione di esperimenti.

Esercizio 50. Continuiamo l'esercizio precedente, restringendo l'attenzione ai soli fattori P ed S. Supponiamo di aver scelto due livelli per ciascuno. Supponiamo di effettuare degli esperimenti e di osservare, per ciascuna combinazione di livelli, i seguenti 3 valori della X (ad esempio gli incassi in tre giornate relative a ciascuna scelta dei livelli):

(P_1, S_1): 3.5, 4, 3.8; (P_2, S_1): 4.2, 4.5, 3.7; (P_1, S_2): 4, 4.2, 4.5; (P_2, S_2): 4.5, 4.3, 4.8.

Stimare i parametri del modello. Discutere intuitivamente gli effetti dei fattori che emergono da questi calcoli. Indicare la combinazione di fattori che sembra più favorevole.

Esercizio 51. Relativamente all'esercizio precedente, supponiamo di voler eseguire ulteriori esperimenti, essendoci accorti che ci sono altri due fattori di disturbo, di cui bisogna contenere l'influenza: l'ordine temporale dei giorni in cui si effettuano gli esperimenti può avere influenza (la notizia degli sconti circola di più al passare dei giorni, ad esempio), e la vicinanza al fine settimana anche. Come si può progettare un piano di esperimenti, per valutare l'effetto dei fattori P ed S, in modo da contenere l'effetto dei due fattori di disturbo sopra citati?

Esercizio 52. Sempre relativamente ai problemi precedenti, concentriamo l'attenzione solo sul fattore di controllo S, di cui continuiamo a considerare due livelli S_1 ed S_2. Facciamo alcuni esperimenti e registriamo i seguenti valori di X per questi due livelli: per S_1: 4.2, 4.5, 3.7; per S_2: 4.5, 4.3, 4.8 (sono i risultati precedenti nelle condizioni P_2). Al 95%, possiamo affermare che il fattore S ha influenza sugli incassi?

Esercizio 53. Con riferimento all'ultimo esercizio, supponiamo di realizzare 4 livelli di sconto, S_j con $j = 1, ..., 4$. Supponiamo di effettuare degli esperimenti e misurare i seguenti incassi: per S_1: 4.2, 4.5, 3.7; per S_2: 4.5, 4.3, 4.8; per S_3: 5.4, 5.5, 5.7; per S_4: 5.5, 5.3, 5.8. Al 95%, possiamo affermare che il fattore S ha influenza sugli incassi?

Esercizio 54. Svolgere gli esercizi 8.1 e 8.4 del libro di testo.

Esercizio 55. Con quale significatività è possibile rigettare l'ipotesi che i seguenti dati

$$-0.156, 2.6, 4.87, -0.267, 3.6, -0.44, 5.61, 2.85, -0.222, 3.06$$

siano un campione gaussiano di rango 10 con media nulla?

Esercizio 56. L'astronomo svizzero R. Wolf (1816-1893) effettuò un esperimento lanciando un dado da gioco 20000 volte ottenendo i seguenti risultati:

Faccia	1	2	3	4	5	6	
Occorrenze	3246	3449	2897	2841	3635	3932	20000

Si confronti il risultato dell'esperimento con l'ipotesi che i possibili risultati del lancio del dado siano equiprobabili. (Si consideri una significatività del 95%)

Esercizio 57. Un campione di 20 individui viene sottoposto a due diverse prove attitudinali che consistono nel portare a termine certe azioni in un tempo determinato. Gli individui vengono classificati a_+ se hanno superato la prima prova, a_- se l'hanno fallita e analogamente b_+ e b_- per la seconda prova. I risultati della sperimentazione sono i seguenti:

	a_+	a_-
b_+	5	1
b_-	4	10

Si può affermare che gli esiti delle due prove siano indipendenti? (Si consideri una significatività del 95%)

Esercizio 58. Si supponga di avere 40 osservazioni sperimentali di una certa quantità X suddivise nel modo seguente: 4 volte X è stato inferiore a 4, 28 volte si è ottenuto $4 \leq X \leq 6$ e le rimanenti 8 volte X era più grande di 6. È ragionevole modellizzare X come una v.a. gaussiana di media 5 e varianza 1? (Si consideri una significatività del 95%)

Esercizio 59. Si consideri il seguente campione di rango 6 di una v.a. X:

$$2.56, 3.27, 3.02, 3.05, 4.28, 3.72.$$

Si può affermare con una significatività del 95% che la media della distribuzione di X è diversa da zero?

Esercizio 60. Venti osservazione di una quantità che può assumere i valori interi da 0 a 4 hanno dato come risultato le seguenti frequenze

Valore	0	1	2	3	4
Occorrenze	5	7	6	2	0

Si può affermare, con una confidenza almeno del 99%, che la quantità corrisponde a una v.a. Binomiale?

1.1 Esercizi supplementari

1) Trovare λ tale che $P(X > \lambda) = 0.7$, con X di tipo $N(-3, 1)$ (usare le tavole). Trovare λ tale che $P(X < \lambda) = 0.1$, con X di tipo $N(5, 1)$.

2) Sia X una v.a. gaussiana di varianza $\sigma^2 = 1$. Sperimentalmente, è stato ottenuto un campione $x_1, ..., x_{20}$ con media $\bar{x} = 3.5$. Determinare una stima della media della v.a. da cui è stato estratto il campione, con l'intervallo di confidenza al 95%.

3) Relativamente all'esercizio 2), come si può ottenere una precisione della stima pari a 0.1, sempre al 95%?

4) Ripetere l'esercizio 2) senza l'ipotesi di conoscere la varianza, avendo invece dal campione il valore $s = 1.2$.

5) Nella stima della media μ di una v.a. X, in cui si ottiene il risultato $\mu = \overline{x} \pm \frac{\sigma q_{1-\frac{\alpha}{2}}}{\sqrt{n}}$, cosa rappresenta α?

6) Su 1000 pezzi esaminati, 5 non corrispondono a determinati standards. Stimare la probabilità di trovare un pezzo fuori regola, al 95%.

7) Su 70 persone interpellate, 25 dichiarano di non acquistare un certo prodotto, le altre sì. Stimare la probabilità che una persona generica acquisti quel prodotto, al 90%.

8) Supponiamo che ad un servizio tramite internet siano iscritte 10.000 persone e che ogni persona si connetta, per usufruire di quel servizio, in media un

giorno ogni dieci. Esprimere (senza svolgere il calcolo completo) la probabilità che nell'arco di una giornata ci siano più di 100 connessioni. Fornire inoltre una formula approssimata utile per il calcolo.

9) Supponiamo che con probabilità 0.00001 un messaggio, inviato in una rete di comunicazione, non arrivi a destinazione. Per aumentare la probabilità che un messaggio venga ricevuto, viene usato il trucco di spedirlo due volte (da considerarsi indipendenti). Qual'è la probabilità che con questo metodo il messaggio venga ricevuto?

10) Una ditta che fornisce pezzi di ricambio deve servire 500 potenziali clienti. Ciascuno si presenta, nell'arco di una giornata, con probabilità 0.02. Come calcolereste la probabilità che si presentino più di 20 clienti nella stessa giornata? Accettando una probabilità 0.001 di non poter soddisfare le richieste giornaliere, quanti pezzi di ricambio devono essere disponibili ogni giorno? L'esercizio non deve essere svolto numericamente, ma solo indicando la strategia ed i ragionamenti.

11) Sia X una v.a. gaussiana di media 5 e varianza $\sigma^2 = 1$. Sperimentalmente, è stato ottenuto un campione $x_1, ..., x_{20}$ con media $\bar{x} = 3.5$. Valutare al 90% se il campione proviene da X.

12) Ripetere l'esercizio 11) senza l'ipotesi di conoscere la varianza, avendo invece dal campione il valore $s = 1.2$.

13) Come decidereste se usare una regressione lineare oppure nonlineare per descrivere l'eventuale legame tra due stringhe di dati $x_1, ..., x_n$ e $y_1, ..., y_n$?

14) Sulla base dei dati che seguono, esprimete un'opinione, basata su un calcolo statistico, circa il fatto se il prezzo di un certo prodotto dipende dal peso del prodotto. Relativamente a cinque esemplari di peso $P = 1$, 2, 2.5, 3.8, il prezzo fatto dalla ditta è risultato di $V = 150\$$, 160\$, 200\$, 300\$. Scegliendo di acquistare un pezzo di lunghezza 1.5 che prezzo vi aspettereste?

15) Supponiamo che la bontà di trasmissione di un segnale venga misurata da due strumenti diversi. Il primo rileva un segnale corretto 95 volte su 100; il secondo solo 85 volte su 100. Questa differenza può essere dovuta a fluttuazioni casuali nei misuratori, oppure al fatto che essi lavorano in modo diverso, quindi producono statistiche diverse. Al 95% ritenete che i due strumenti di misura lavorino nello stesso modo?

16) Valutare al 98% se un certo campione $x_1, ..., x_{15}$, con $\bar{x} = 7.3$, $s^2 = 0.8$, può essere stato estratto da una distribuzione gaussiana di media 7.

17) Supponiamo che X abbia la proprietà che $\log_{10} X$ è di tipo $N(1, 2)$ (X è log-normale). Trovare λ tale che $P(X < \lambda) = 0.2$.

18) Sia X una v.a. gaussiana di varianza $\sigma^2 = 7$. Sperimentalmente, è stato ottenuto un campione $x_1, ..., x_{50}$ con media $\bar{x} = 25$. Determinare, per la stima della media, l'intervallo di confidenza al 99%. Come si può ottenere una precisione della stima pari a 0.01 al 99%?

19) Se un sensore ha probabilità 0.1 di non accorgersi del segnale che deve ricevere, qual'è la probabilità che almeno uno tra due sensori in funzione si accorga del segnale?

20) Determinare il numero M di sportelli che un ufficio deve avere per non far attendere alcun utente con probabilità 0.99, sapendo i seguenti dati: esso

serve una popolazione di 1000 utenti, ciascuno ha probabilità 1/500 di essere presente (ad un generico istante). Indicare solamente il procedimento per calcolare M, senza svolgere i calcoli. E' preferibile una risoluzione interamente basata sull'approssimazione di Poisson.

21) Relativamente alle concentrazioni 0.1, 0.3, 0.5, 0.7, 0.9 di una certa sostanza, si è rilevato che il filato ha spessore 2.5, 2.8, 3.2, 3.8, 4.8. Si vuole calcolare la concentrazione che produce uno spessore pari a 4. Come procedete? Indicate anche eventuali considerazioni alternative che decidete di non svolgere per limiti di calcolo.

22) Stabilire, al 95%, se la media di un prodotto, inizialmente pari a 0, è peggiorata, sulla base di un campione di numerosità 30, con media sperimentale 0.3, e deviazione standard sperimentale 0.6. Si supponga di sapere a priori che la media può solo essere aumentata. Calcolare la potenza di questo test (usando l'approssimazione gaussiana), per una deviazione della media di 0.5.

23) Lo scorso anno il 70% delle persone acquistava il prodotto A, le restanti il B. Indagini di quest'anno su N persone dicono che solo il 60% di esse acquista il prodotto A. Al 95%, si tratta di una fluttuazione casuale o di un cambio di tendenza dei gusti? Impostare la risoluzione, senza svolgere necessariamente tutti i calcoli, sia nel caso in cui $N = 10$, sia per $N = 1000$.

24) Calcolare la probabilità di vendere un sono pezzo, o nessuno, sapendo di avere 50 potenziali clienti, ciascuno con probabilità 0.1 di comprare il prodotto. Usare una formula esatta ed anche una approssimata. Calcolare anche il numero medio di pezzi che si prevede usualmente di vendere.

25) In un gioco in cui si lanciano due dadi, si vince se almeno uno produce il valore 6. Calcolare la probabilità di vincere.

26) State producendo un nuovo prodotto, con caratteristiche ancora incognite. Volete determinare la media di una caratteristica con precisione 0.1, correndo un rischio 0.05 di dare una stima sbagliata. Spiegate cosa fate.

27) Si prevede che due persone su dieci trascorrano vacanze all'estero. Dopo una piccola indagine effettuata su 15 persone, si rileva che solo tre di esse vanno all'estero. Stabilire al 90% se la previsione era errata.

2 Alcune soluzioni

Riportiamo la risoluzione (eventualmente solo il risultato o lo schema di risoluzione) di alcuni degli esercizi del primo elenco. Altri dettagli di risoluzione si possono trovare sul libro di testo, nella trattazione degli esempi da cui questi esercizi sono stati estratti.

1: $\frac{5\cdot 4}{2} : \frac{90\cdot 89}{2}$. **2**: $\binom{n}{k}\frac{1}{2^n}$. **3**: $\binom{n-1}{k-1}\frac{1}{2^n}$. **5**: $0.95+0.1\cdot 0.05$. **6**. Un sistema in parallelo non funziona se entrambe le componenti non funzionano. Consideriamo gli eventi $A =$ la componente c_1 non funziona, $B =$ la componente c_1 non funziona. L'evento $A\cap B$ è $=$ il sistema non funziona. Essendo A e B indipendenti, vale $P(A\cap B) = P(A)\cdot P(B)$. Sappiamo poi che $P(A) = 1-p_1$ e $P(B) = 1-p_2$, quindi la probabilità che il sistema non funzioni è $(1-p_1)\cdot(1-p_2)$.

7. P(poter raggiungere la destinazione) $= 1 - P$(entrambe le strade interrotte) $= 1 - [P$(prima strada interrotta) $\cdot P$(seconda strada interrotta)$] = 1 - (0.01)^2 = 0.9999$.

8. Un sistema in serie funziona se entrambe le componenti funzionano. Quindi P(il sistema funzioni) $= P(A \cap B)$ dove $A =$ la prima componente funziona, $B =$ la seconda componente funziona. Essendo indipendenti vale P(il sistema funzioni) $= P(A) \cdot P(B) = (0.9)^2 = 0.81$.

9. $1 - 0.902\,5$.

10. Indichiamo con X_1, ..., X_{1000} i risultati degli esami: X_1 vale 1 se il primo pezzo esaminato è rotto, 0 altrimenti, e così via. Le X_i sono quindi v.a. di Bernoulli di parametro 0.01, che supponiamo indipendenti. La v.a. $X = X_1 + ... + X_{1000}$ conta i pezzi rotti, ed è una $B(1000, 0.01)$. La probabilità di trovare 5 pezzi rotti su 1000 esaminati vale pertanto

$$P(X = 5) = \binom{1000}{5} (0.01)^5 (0.99)^{995}$$

che vale $3.745\,3 \times 10^{-2}$. Usando poi il teorema degli eventi rari si approssima la precedente quantità con $e^{-\lambda} \frac{\lambda^5}{5!}$ prendendo $\lambda = np = 10$, quindi $e^{-\lambda} \frac{\lambda^k}{5!} = e^{-10} \frac{10^5}{5!} = 3.783\,3 \times 10^{-2}$.

14. $\binom{n}{k} p^k (1-p)^{n-k}$ per $k = 0, 1, ..., n$, 0 altrimenti. **15**. p, $p(1-p)$, np, $np(1-p)$. **16**. $P(A \cap \overline{B}) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(\overline{B})$. **17**. Usare la strategia del test di Fisher. **20**. Di nuovo col testi di Fisher. La probabilità cercata è 0.29371, ottenuta come

$$\frac{\binom{6}{1}\binom{9}{4}}{\binom{15}{5}} + \frac{\binom{6}{0}\binom{9}{5}}{\binom{15}{5}}.$$

21. Si applica un test binomiale, ovvero si calcola la coda di probabilità P(2 su 10) + P(1 su 10) + P(0 su 10) usando la binomiale con $p = 0.3$. Si trova il valore 0.382. Questa è la probabilità che, pur essendo 30% la probabilità di vendita di quel modello, su 10 acquirenti due o meno di due lo scelgano. Siccome questa probabilità non è piccola, non possiamo rifiutare l'ipotesi che sia 30% la probabilità di vendita di quel modello. **22**. Il sistema funziona se nessun relais è bloccato e contemporaneamente almeno uno non è ossidato. Il risultato finale è $0.991 \cdot \left(1 - 3.43 \cdot 10^{-7}\right)$. **23**. $\frac{\lambda}{\lambda - t}$, $\frac{1}{\lambda}$, $\frac{1}{\lambda^2}$. **25**. $n\mu$ (non serve l'indipendenza); $n\sigma^2$ (usando l'indipendenza). **27**. Detta X_k la v.a. che vale 1 se si ha successo nella k-esima prova, posto $\mu = p = 0.1$ (media di X_k) e $\sigma^2 = p(1-p)$ (varianza di X_k), abbiamo

$$\begin{aligned} P(X_1 + ... + X_{500} \leq 45) &= P\left(\frac{X_1 + ... + X_{500} - 500 \cdot \mu}{\sigma\sqrt{500}} \leq \frac{45 - 500 \cdot \mu}{\sigma\sqrt{500}}\right) \\ &\sim P(Z \leq -0.745\,35) = F(-.745\,35) = 1 - F(0.745\,35) = 1 - 0.77 = 0.23 \end{aligned}$$

dove l'approssimazione è dovuta al teorema limite centrale, e dove F è la funzione di distribuzione della $N(0,1)$. **28**. $R^3 + 3(1-R)R^2$. **29**. $L = 0.000512$.

30. Conviene calcolare prima $F_Y(y) = P(Y \leq y)$ e poi calcolare $f_Y(y)$ come derivata di $F_Y(y)$. In questo esempio vale

$$\begin{aligned} P(Y \leq y) &= P(|X| \leq y) = P(-y \leq X \leq y) \\ &= F_X(y) - F_X(-y). \end{aligned}$$

Non importa conoscere la forma esplicita di questa funzione. Infatti possiamo derivare (ricordando che $f_X(y) = \frac{d}{dy} F_X(y)$):

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}\left[F_X(y) - F_X(-y)\right] = f_X(y) - f_X(-y)(-1) \\ &= f_X(y) + f_X(-y) = \frac{2}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}. \end{aligned}$$

33. E' il modello con media $\overline{x} = 5.15$ e deviazione standard $S = 0.288$. **34**. Usando il metodo dei momenti, $10p = \overline{x} = 1.14$, quindi $p = 0.114$. **35**. $\lambda = 1.14$. **36**. Usando il metodo dei momenti, $\frac{1}{\lambda} = \overline{x} = 4.8$, quindi $\lambda = 0.208$. **37**. $\mu = 21.83 \pm 1.6$. **38**. $0.5 \geq \delta = \frac{\sigma q_{1-\frac{\alpha}{2}}}{\sqrt{n}}$, da cui $n = 62$. **39**. $S = 2.45$, da cui $\mu = 21.83 \pm 1.96$. **40**. $\mu = 420.87 \pm 12.337$. $n = 49$. **42**. Chiamiamo X la v.a. da cui è estratto il campione. L'ipotesi nulla è $H_0 = \{E(X) = 500\}$ mentre l'ipotesi alternativa è $H_1 = \{E(X) > 500\}$. Si sceglie quindi una strategia di test unilaterale e si adotta un test t di Student. Calcoliamo $t = \sqrt{10}\frac{508-500}{5} = 5.0596$. Nell'ipotesi H_0, questo è un valore estratto da una t di Student $T(9)$ a 9 gradi di libertà. Pertanto la regione di rigetto (vista l'ipotesi H_1 unilaterale) è l'intervallo $\left(t^9_{0.95}, \infty\right)$ dove $t^9_{0.95}$ indica il quantile della $T(9)$ relativo ad un'area pari a 0.95. Vale $t^9_{0.95} = 1.833$ (dalle tavole). Siccome $5.0596 > 1.833$, rifiutiamo l'ipotesi H_0 (ricordiamo che nell'ipotesi H_0 la probabilità che fosse $T(9) > t^9_{0.95}$ era 0.05, il livello da noi considerato "impossibile" in base alla nostra scelta di significatività). Quindi, al livello fissato di significatività, riteniamo che il nuovo metodo di produzione migliori il prodotto. **43**. Innanzi tutto modifichiamo l'esecuzione del test in accordo col fatto che la varianza è nota. A partire dai dati calcoliamo ora la grandezza $z = \sqrt{10}\frac{508-500}{4} = 6.3246$, che, nell'ipotesi H_0, è un valore estratto dalla v.a. $Z = \sqrt{10}\frac{\overline{X}-500}{4}$, che è una $N(0,1)$. Va quindi confrontato col quantile gaussiano $q_{0,95} = 1,645$. Il test risulta significativo. Fatte queste premesse, valutiamo la potenza di questo test di accorgersi di una nuova media pari a 505. Introduciamo la nuova ipotesi $H_2 = \{E(X) = 505\}$. L'errore di seconda specie è la probabilità che il test abbia esito non significativo (ovvero che si trovi $Z \leq q_{0,95}$) nell'ipotesi che la distribuzione di provenienza fosse una $N(505, 16)$. Qui è sempre $Z = \sqrt{10}\frac{\overline{X}-500}{4}$, che però sotto l'ipotesi H_2 non è più $N(0,1)$. Dobbiamo quindi prima di tutto trovare la legge di Z. Ricordando che $E\left[\overline{X}\right] = E[X] = 505$, $Var\left[\overline{X}\right] = \frac{16}{10} = 1,6$, Si trova $E[Z] = \sqrt{10}\frac{E\left[\overline{X}\right]-500}{4} = 3.9528$, $Var[Z] = 1$ (verificare). Quindi Z è una $N(3.9528; 1)$. L'errore di seconda specie è quindi uguale a

$$\begin{aligned} P(Z \leq q_{0,95}) &= F(1.645 - 3.9528) = F(-2.3078) \\ &= 1 - F(2.3078) = 1 - 0.989. \end{aligned}$$

La potenza è quindi 0.989. Per rilevare simili variazioni di media questo risulta essere un test molto potente. **46**. $\overline{x} = 2.5$, $\overline{y} = 4.5$, $\sigma_X \sim \sqrt{\frac{1}{n}\sum(x_i - \overline{x})^2} = 1.118$, $\sigma_Y \sim \sqrt{\frac{1}{n}\sum(y_i - \overline{y})^2} = 2.29$, $Cov(X,Y) \sim \frac{1}{n}\sum(x_i - \overline{x})(y_i - \overline{y}) = 2.5$, e quindi il coefficiente di correlazione $\rho(X,Y) \sim 0.9759$. Il legame lineare è accettabile. Risulterà $a > 0$. Le stime dei parametri sono $a \sim \frac{2.5}{1.25} = 2$, $b = 4.5 - 2 \cdot 2.5 = -0.5$. La retta di regressione è $Y = 2 \cdot X - 0.5$. Vale inoltre $\sigma_\varepsilon^2 = \sigma_Y^2\left(1 - \rho^2\right) = 0.249$. Per $X = 5$ prevediamo $Y = 9.5 \pm \delta$ al 95%, dove δ è quel numero tale che $P(|Y(5) - 9.5| > \delta) = 0.05$. Qui $Y(5)$ indica la v.a. Y in corrispondenza del valore $X = 5$, ovvero la v.a. $Y(5) = 9.5 + \varepsilon$, dove ε è una $N(0; 0.249)$. Dobbiamo quindi trovare δ tale che $P(|\varepsilon| > \delta) = 0.05$, ovvero $P(\varepsilon < \delta) = 0.975$, ovvero $F\left(\frac{\delta}{0.499}\right) = 0.975$, ovvero $\frac{\delta}{0.499} = q_{0.975} = 1.96$, e quindi infine $\delta = 0.978$. **49**. Supponiamo di suddividere l'entità o qualità della pubblicità in n_P livelli P_i, di fare lo stesso per n_S livelli degli sconti S_j, e per n_Q livelli della qualità Q_k della merce. Fissata una combinazione (P_i, S_j, Q_k) dei fattori, l'incasso sarà descritto da una v.a. X_{ijk} gaussiana di media μ_{ijk}. Decomponiamo poi questa media in componenti corrispondenti agli effetti dei singoli fattori e delle loro combinazioni, pervenendo al modello

$$X_{ijk}^{(r)} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijk}^{(r)}.$$

L'indice $r = 1, ..., R_{ijk}$ descrive, fissati ijk, i diversi esperimenti che verranno effettuati relativamente alla combinazione (P_i, S_j, Q_k). Le v.a. $\varepsilon_{ijk}^{(r)}$ si suppongono indipendenti, tutte $N(0, \sigma^2)$. Il significato dei vari parametri è ad esempio: $\overline{\mu}$ è la media complessiva al variare di tutti i livelli di tutti i fattori; $\alpha_i = \mu_{i\cdot\cdot} - \mu$ è l'effetto del fattore P al livello P_i, dove abbiamo posto $\mu_{i\cdot\cdot}$ uguale alla media di P_i al variare di tutti i livelli dei fattori S e Q; e così via (il significato di $(\alpha\beta)_{ij}$, ad esempio, è spiegato nella soluzione dell'esercizio seguente). **50**. La media aritmetica complessiva è

$$\overline{x} = \frac{3.5 + 4 + 3.8 + 4.2 + 4.5 + 3.7 + 4 + 4.2 + 4.5 + 4.5 + 4.3 + 4.8}{12} = 4.1667.$$

Questa è la stima di μ. La media di P_1 al variare di tutti i livelli del fattore S è

$$\overline{x}_{1\cdot} = \frac{3.5 + 4 + 3.8 + 4 + 4.2 + 4.5}{6} = 4.0$$

e questa è la stima di $\mu_{1\cdot}$, per cui la stima di α_1 è $\widehat{\alpha}_1 = \overline{x}_{1\cdot} - \overline{x} = -0.166$. La media di P_2 al variare di tutti i livelli del fattore S è

$$\overline{x}_{2\cdot} = \frac{4.2 + 4.5 + 3.7 + 4.5 + 4.3 + 4.8}{6} = 4.333$$

e questa è la stima di $\mu_{2\cdot}$, per cui la stima di α_2 è $\widehat{\alpha}_2 = \overline{x}_{2\cdot} - \overline{x} = 0.166$. Così di seguito vale $\overline{x}_{\cdot 1} = 3.95$, $\overline{x}_{\cdot 2} = 4.383$, $\widehat{\beta}_1 = \overline{x}_{\cdot 1} - \overline{x} = -0.216$, $\widehat{\beta}_2 = \overline{x}_{\cdot 2} - \overline{x} = 0.216$. Infine, $\left(\widehat{\alpha\beta}\right)_{11} = (\overline{x}_{11} - \overline{x}_{\cdot 1}) - (\overline{x}_{1\cdot} - \overline{x})$, quindi, essendo $\overline{x}_{11} = \frac{3.5+4+3.8}{3} = 3.766$, vale $\left(\widehat{\alpha\beta}\right)_{11} = -0.016$. Gli altri $\left(\widehat{\alpha\beta}\right)_{ij}$ si calcolano analogamente. Molto intuitivamente, vediamo che il fattore P ha un po' meno influenza del fattore S (i valori $\widehat{\beta}_j$ sono un po' più grandi degli $\widehat{\alpha}_i$), mentre l'effetto incrociato dei due fattori, almeno per quanto emerge dalla combinazione $(\alpha\beta)_{11}$, è praticamente inesistente. Confrontando infine le medie relative alle singole combinazioni di fattori, ovvero $\overline{x}_{11} = 3.766$, $\overline{x}_{21} = 4.133$, $\overline{x}_{12} = 4.233$, $\overline{x}_{22} = 4.533$, vediamo che la combinazione (P_2, S_2) sembra la più conveniente. **51**. Abbiamo 4 combinazioni (P_i, S_j) dei fattori di controllo, che possiamo indicare con a, b, c, d. Consideriamo allora 4 livelli per il fattore "vicinanza al fine settimana" e 4 livelli per "l'ordine dei giorni di prova". Possiamo ora prendere un quadrato latino 4×4. **52**. Si tratta di confrontare due gruppi, provenienti da gaussiane, per capire se provnegono dalla stessa gaussiana oppure no (in quest'ultimo caso S avrà influenza sugli incassi). Possiamo applicare un test t di Student. Il test per campioni indipendenti è più potente ed è consigliato se lo si conosce. Svolgiamo l'esercizio usando il più semplice test per dati appaiati, che però è meno potente. Consideriamo allora il campione x_1, x_2, x_3 ottenuto facendo le differenze (appaiamo i dati nel modo ovvio, anche se artificiale essendo in realtà i campioni del tutto indipendenti): 0.3, -0.2, 1.1. Vale $\overline{x} = \frac{0.3-0.2+1.1}{3} = 0.4$, $S^2 = \frac{1}{2}\sum_{i=1}^{3}(x_i - \overline{x})^2 = 0.43$, $S = 0.655$, quindi $t = \sqrt{3}\frac{\overline{x}}{S} = 1.0577$ (l'ipotesi nulla è $\mu = 0$). Questo va confrontato col quantile t di Student a 2 gradi di libertà $t^2_{0.975}$ (svolgiamo un test a due code), che vale 4.303. Il test non risulta significativo. Con i dati a disposizione e con la significatività richiesta, non possiamo affermare che S ha un influsso sugli incassi. **53**. Effettuiamo un'analisi della varianza ad una via.