

Esercizi di Statistica Medica

Materiale di supporto ai corsi per le Professioni Sanitarie – Università di Padova

Corrado Lanera

Table of contents

Introduzione	3
Come utilizzare questo materiale	3
Suggerimenti per lo studio	4
Aggiornamenti	4
1 Statistica descrittiva 1	5
1.1 Domande	5
2 Probabilità - Teorema della somma	7
2.1 VALUTAZIONE DEL TASSO DI POSITIVITÀ COVID-19 NELL'OSPEDALE	7
2.2 Domande	7
3 Probabilità - Condizionamento	8
3.1 Domande	8
4 Probabilità - Valori Predittivi	9
4.1 Contesto A	9
4.2 Domande A	9
4.3 Contesto B	9
4.4 Domande B	10
References	11

Introduzione

Questo documento raccoglie una selezione di esercizi di **Statistica Medica** pensati per accompagnare lo studio teorico e le lezioni frontali dei corsi tenuti da Corrado Lanera e destinati alle *Professioni Sanitarie* dell'Università di Padova.

Gli esercizi coprono i principali argomenti affrontati nel corso:

- **Statistica descrittiva**, con esempi di sintesi numerica e grafica dei dati.
- **Concetti fondamentali di probabilità**, utili a comprendere il ragionamento diagnostico.
- **Probabilità condizionata e teorema di Bayes**, applicati alla valutazione delle prestazioni dei test diagnostici.
- **Distribuzioni di probabilità**, per modellare variabili casuali discrete (Binomiale e Poisson) e continue (Normale e Normale Standard).
- **Stima puntuale e intervalli di confidenza**, per quantificare l'incertezza delle stime.
- **Test di ipotesi** per misure continue, per prendere decisioni basate sui dati osservati.

Ogni esercizio è pensato per favorire la comprensione intuitiva dei concetti, più che la semplice applicazione di formule. L'obiettivo è abituarsi a **leggere, interpretare e ragionare sui dati** in contesti clinici e di ricerca.

Gran parte degli esercizi sono tratti da testi di riferimento, articoli scientifici o dataset reali. Si veda la bibliografia per i riferimenti completi.

Come utilizzare questo materiale

Il materiale può essere usato in due modi complementari:

1. **In parallelo alle lezioni**, per consolidare i concetti appena introdotti.
2. **Come preparazione all'esame**, selezionando gli argomenti di maggiore difficoltà o interesse.

Ogni sezione presenta brevi richiami teorici, seguiti da esercizi guidati e talvolta da domande aperte per stimolare la discussione.

Alcuni esercizi fanno riferimento a **dataset reali o simulati** usati a lezione.

Suggerimenti per lo studio

- Leggi sempre **il testo dell'esercizio** con attenzione prima di passare ai calcoli.
 - Cerca di **interpretare il significato clinico** dei risultati numerici: la statistica non è solo calcolo, ma anche comprensione del contesto.
 - Confronta i tuoi risultati con quelli discussi a lezione; soprattutto, discutli con i colleghi.
 - Se utilizzi R, prova a **riprodurre le analisi passo per passo**, verificando come cambiano i risultati al variare dei dati.
-

Aggiornamenti

Questo materiale è in **costante aggiornamento**.

Eventuali correzioni, estensioni o nuove versioni saranno rese disponibili attraverso i canali ufficiali del corso o tramite il profilo docente.

1 Statistica descrittiva 1

I dati riportati nella tabella qui sotto sono un campione di un insieme più ampio di dati raccolti su persone dimesse da un ospedale selezionato della Pennsylvania nell'ambito di una revisione retrospettiva delle cartelle cliniche sull'uso degli antibiotici negli ospedali.

ID no.	Duration of hospital stay	Age	Sex 1 = M 2 = F	First temp. following admission	First WBC ($\times 10^3$) following admission	Received antibiotic? 1 = yes 2 = no	Received bacterial culture? 1 = yes 2 = no	Service 1 = med. 2 = surg.
1	5	30	2	99.0	8	2	2	1
2	10	73	2	98.0	5	2	1	1
3	6	40	2	99.0	12	2	2	2
4	11	47	2	98.2	4	2	2	2
5	5	25	2	98.5	11	2	2	2
6	14	82	1	96.8	6	1	2	2
7	30	60	1	99.5	8	1	1	1
8	11	56	2	98.6	7	2	2	1
9	17	43	2	98.0	7	2	2	1
10	3	50	1	98.0	12	2	1	2
11	9	59	2	97.6	7	2	1	1
12	3	4	1	97.8	3	2	2	2
13	8	22	2	99.5	11	1	2	2
14	8	33	2	98.4	14	1	1	2
15	5	20	2	98.4	11	2	1	2
16	5	32	1	99.0	9	2	2	2
17	7	36	1	99.2	6	1	2	2
18	4	69	1	98.0	6	2	2	2
19	3	47	1	97.0	5	1	2	1
20	7	22	1	98.2	6	2	2	2
21	9	11	1	98.2	10	2	2	2
22	11	19	1	98.6	14	1	2	2
23	11	67	2	97.6	4	2	2	1
24	9	43	2	98.6	5	2	2	2
25	4	41	2	98.0	5	2	2	1

1.1 Domande

1. Calcolare media, moda e mediana della durata di ospedalizzazione.

2. Calcolare il varianza, deviazione standard, coefficiente di variazione, range massimale e range interquartile (differenza tra il III e il IV quartile) della durata di ospedalizzazione. Descrivere a parole in linguaggio naturale tali dati sulla base delle informazioni appena ricavate (un paragrafo).
3. I tempi riportati sono in giorni, supponiamo di volerli esprimere in ore, ovvero moltiplicandoli tutti per 24; che effetto avrebbe questa modifica nei valori di media, moda, mediana, varianza, deviazione standard, coefficiente di variazione, range massimale e range interquartile? Verificare la risposta eseguendo i conti.
4. Quali delle variabili riportate possono essere considerate continue, quali categoriche? Tra le categoriche, come sarebbe opportuno considerare Sex e Service (binarie, nominali o ordinali)? Motivare le risposte.
5. Possiamo dire che una riga della tabella rappresenta una unità statistica? Sì, no, perché?
6. Disegnare (a mano) un grafico a barre per le frequenze della variabile Sex, e un box-plot per il tempo di ospedalizzazione. Cosa si può dire guardando questi due grafici? (un paragrafo ciascuno)
7. Disegnare due box-plot sovrapposti/affiancati in uno stesso grafico (con un unico asse x (se orizzontale) o y (se verticali, scelta a piacere)) per i tempi di ospedalizzazione, separati per maschi e femmine. Cosa si può dire guardando questo grafico? (un paragrafo)

(Townsend et al. 1979)

2 Probabilità - Teorema della somma

2.1 VALUTAZIONE DEL TASSO DI POSITIVITÀ COVID-19 NELL'OSPEDALE

Sei un membro del team di controllo infezioni di un grande ospedale e ti viene chiesto di stimare il tasso complessivo di positività al COVID-19 tra i pazienti ricoverati. Non tutti i reparti hanno la stessa incidenza di casi, a causa delle diverse tipologie di pazienti e procedure.

Per facilitare l'analisi, l'ospedale è stato suddiviso in cinque macro-reparti principali, per i quali sono disponibili i seguenti dati:

Reparto	Porzione pazienti totali	Tasso positività CiViD-19 nel reparto
Medicina Interna	35% (0.35)	20% (0.20)
Chirurgia	25% (0.25)	8% (0.08)
Pediatria	15% (0.15)	12% (0.12)
Terapie Intensive	10% (0.10)	40% (0.40)
Altri Reparti	15% (0.15)	5% (0.05)

2.2 Domande

1. Calcolare il tasso di positività complessivo al COVID-19 per l'intero ospedale, utilizzando il Teorema della Probabilità Totale.
2. Qual è la probabilità che un paziente scelto a caso nell'ospedale sia positivo al COVID-19?

3 Probabilità - Condizionamento

Uno studio basato sui dati raccolti dal Registro Medico delle Nascite della Norvegia ha esaminato i tassi di fertilità in base agli esiti di sopravvivenza delle nascite precedenti [1]. I dati sono presentati nella Tabella seguente.

	First birth	Continuing to second birth	Second birth outcome	Continuing to third birth	Third birth outcome
Perinatal outcome	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>
D	7022	5924	D 368	277	D 39
					L 238
			L 5556	3916	D 115
					L 3801
L	350,693	265,701	D 3188	2444	D 140
					L 2304
			L 262,513	79,450	D 1005
					L 78,445

Note: D = dead, L = alive at birth and for at least one week.

3.1 Domande

1. Qual è la probabilità di avere un parto vivo (L) in occasione di una seconda nascita, ipotizzando che l'esito della prima gravidanza sia stato un parto morto (D)?
2. Rispondere al Problema 1. ipotizzando l'esito della prima gravidanza sia stato un parto vivo.
3. Qual è la probabilità di 0, 1 e 2+ gravidanze aggiuntive se il primo parto è stato un parto vivo.
4. Rispondere al Problema 3. ipotizzando il primo parto sia stato un parto morto.

(SKJÆRVEN et al. 1988)

4 Probabilità - Valori Predittivi

4.1 Contesto A

È stato condotto uno studio per valutare l'accuratezza dell'esposizione autodichiarata al fumo di sigaretta in utero. È stato effettuato un confronto tra le dichiarazioni delle figlie sul fumo delle madri durante la gravidanza e le autodichiarazioni delle madri sul proprio fumo durante la gravidanza delle figlie. I risultati sono riportati nella Tabella seguente.

Daughter's report of fetal smoke exposure	Mother's report of smoking during pregnancy	<i>N</i>
yes	yes	6685
yes	no	1126
no	yes	1222
no	no	23,227

4.2 Domande A

1. Se l'auto-rapporto della madre è considerato completamente accurata, allora qual è il PV+ del resoconto della figlia, in cui il positivo indica che fuma e il negativo indica che non fuma?
2. Se l'autodichiarazione di una madre è considerata completamente accurata, qual è il PV- del resoconto della figlia?

4.3 Contesto B

Ulteriori dati sul fumo auto-riferito indicano che la madre non è sempre completamente accurata. La cotinina salivare è un marcatore biochimico che, se elevato, è un'indicazione

accurata al 100% del fumo recente.

Supponiamo che se la madre dichiara di non essere fumatrice durante la gravidanza, la cotinina salivare risulti elevata nel 5% dei casi, mentre se la madre dichiara di essere fumatrice durante la gravidanza, la cotinina salivare è elevata il 97% delle volte. Si supponga inoltre che la relazione della figlia non aggiunga ulteriori informazioni sulla probabilità di un livello elevato di cotinina una volta nota l'autodichiarazione della madre.

4.4 Domande B

3. Qual è la probabilità che il livello di cotinina nella saliva di una madre sia elevato durante la gravidanza se la figlia riferisce che la madre fumava in gravidanza?
4. Qual è la probabilità che il livello di cotinina nella saliva della madre non sia elevato durante la gravidanza se la figlia riferisce che la madre fumava in gravidanza?

References

- SKJÆRVEN, ROLV, ALLEN J. WILCOX, ROLV T. LIE, and LORENTZ M. IRGENS. 1988. “SELECTIVE FERTILITY AND THE DISTORTION OF PERINATAL MORTALITY.” *American Journal of Epidemiology* 128 (6): 1352–63. <https://doi.org/10.1093/oxfordjournals.aje.a115088>.
- Townsend, T. R., M. Shapiro, B. Rosner, and E. H. Kass. 1979. “Use of Antimicrobial Drugs in General Hospitals. I. Description of Population and Definition of Methods.” *Journal of Infectious Diseases* 139 (6): 688–97. <https://doi.org/10.1093/infdis/139.6.688>.