

Studio inferenziale sulla relazione tra dolore al collo e l'utilizzo dello smartphone

Riccardo Santi

Sep 2025

Chapter 1

Introduzione

1.1 Panoramica

Questo elaborato fa parte del più ampio lavoro di tesi triennale del mio collega Tommaso Sartorato, laureando in Fisioterapia presso l'Università di Padova. La tesi indaga e approfondisce come l'uso quotidiano dello smartphone influenzi la salute del rachide cervicale. Il mio contributo consiste nell'analisi inferenziale presente nell'elaborato.

I dati che verranno analizzati sono stati raccolti tramite questionario online, tramite canali popolati prevalentemente da studenti universitari di Fisioterapia.

1.2 Domande di ricerca

Il tesista ha individuato come interessanti domande di ricerca l'esistenza o meno delle seguenti relazioni:

1. Relazione tra la presenza del dolore al collo e l'età.
2. Relazione tra la frequenza del dolore al collo e
 - l'età.
 - le ore di utilizzo del proprio smartphone e/o tablet.
 - l'altezza in cui è posto il dispositivo.
 - la frequenza con cui si pratica attività sportiva.
 - le ore passate a studiare e/o al pc.
3. Relazione tra l'intensità del dolore e
 - l'età.
 - le ore di utilizzo del proprio smartphone e/o tablet.
 - l'altezza in cui è posto il dispositivo.
 - le ore passate a studiare e/o al pc.

Chapter 2

Approccio e limiti dello studio

2.1 Approccio allo studio

Siamo interessati alle possibili relazioni causa-effetto tra le variabili rilevate e le variabili risposta riguardanti il dolore cervicale. Data la modalità di raccolta dei dati, questo è un classico esempio di problema di inferenza causale in uno studio osservazionale. In studi del genere, è importante considerare insieme alle variabili responsabili della relazione causa-effetto tutte quelle altre variabili che siano correlate sia con la variabile attribuita alla causa che quella legata all'effetto (si veda ad esempio il capitolo 3 di Wooldridge, 2016, per approfondimenti). Senza questa accortezza gli stimatori utilizzati nello studio risultano distorti, e saggiano semplicemente l'ipotesi che ci sia un'associazione tra variabili, invece di testare l'esistenza di un rapporto di causalità, domanda decisamente più interessante. Per le ragioni sopra esposte, si evitano test marginali (ad esempio, il test del χ^2 di indipendenza o modelli lineari semplici) poiché tali test confrontano solamente distribuzioni marginali, non controllando per potenziali variabili di confondimento e ottenendo di conseguenza una distorsione sistematica.

Verranno invece utilizzati modelli statistici, con opportune covariate elencate in seguito. L'effetto delle variabili sulla risposta di interesse sarà indicato dai coefficienti stimati dai modelli statistici. Tali coefficienti vanno interpretati secondo la concezione *ceteris paribus*, ovvero la variazione della risposta legata al cambiamento di una unità della covariata presa in considerazione. Risponderemo al quesito sull'esistenza di tale relazione grazie ad un semplice test Z di Wald o ad un test del rapporto di verosimiglianza, che testa l'ipotesi che il coefficiente legato ad una covariata sia uguale a 0. Minore sarà il p-value associato al test maggiore sarà l'evidenza empirica che il coefficiente è diverso da 0, e che ci sia quindi una relazione con la risposta. Convenzionalmente si attribuisce una associazione significativa per p-value minori a 0.05, ma non è una regola rigida.

Si precisa che non si stima un unico modello che includa tutte le variabili concomitanti di interesse nelle varie domande di ricerca; questo perché fare ciò aumenterebbe sostanzialmente la varianza delle stime dei singoli coefficienti.

2.2 Alcuni limiti dello studio

Tra i limiti dello studio si evidenziano:

- La scarsa numerosità campionaria, che limita la potenza dei test e la capacità di rilevare effetti di

piccola entità.

- Alcuni intervalli di variabili (ad esempio l'età) sono molto ristretti. Sebbene i modelli impiegati possano fornire predizioni e conclusioni per valori esterni all'intervallo osservato, l'estrapolazione al di fuori dell'intervallo è in Statistica una pratica rischiosa: le conclusioni sono affidabili solo per soggetti con caratteristiche quantomeno simili a quelle presenti nel campione.
- Il campione di risposte ottenuto potrebbe non essere rappresentativo della popolazione. Ad esempio, la maggior parte dei rispondenti, anche se non è stato rilevato, è probabilmente universitaria e consegue lo stesso titolo di studio, data la modalità in cui sono state raccolte le risposte. Se fattori del genere influiscono sulla variabile d'interesse solamente attraverso altre variabili da noi rilevate questo non è un problema, siccome le stime risultano non distorte. Invece, se i fattori non osservati influiscono sulla variabile d'interesse anche al netto delle variabili rilevate le stime risultano distorte e i risultati parzialmente compromessi.
- Ci sono poche variabili rilevate. Probabilmente sarebbe opportuno rilevare altre variabili, ci sono molte altre variabili che andrebbero controllate per affermare che stiamo studiando la reale presenza di causalità. Pensiamo quindi che le nostre stime degli effetti siano sostanzialmente distorte, tuttavia è difficile ottenere dati migliori senza un reale budget. Comunque controllare per le variabili in nostro possesso è sicuramente meglio di limitarsi a testare associazioni marginali.
- Ci sono delle risposte mancanti per alcune osservazioni, per le quali è stata eseguita una semplice imputazione con un modello lineare. In altri casi le risposte erano parzialmente contraddittorie o c'erano evidenti errori di battitura. Dunque, è stata effettuata una pulizia preliminare del dataset.

Chapter 3

Modelli e risultati

3.1 Presenza del dolore cervicale

3.1.1 Presenza del dolore cervicale e età

È stato stimato un modello logit binomiale per la risposta dicotomica al quesito "Indica se hai mai avuto dolore nella regione del collo (cervicale) negli ultimi mesi". Sono stati inclusi nel modello le variabili di controllo sesso, altezza(quadratica) e peso, oltre a età, oggetto dello studio.

Il modello è:

$$\text{logit}(P(\text{Dolore_cervicale} = 1)) = \beta_0 + \beta_1 \cdot \text{sessoM} + \beta_2 \cdot \text{altezza} + \beta_3 \cdot \text{altezza}^2 + \beta_4 \cdot \text{peso} + \beta_5 \cdot \text{età} \quad (3.1)$$

dove $\text{logit}(p) = \ln\left(\frac{p}{1-p}\right)$.

Le stime vengono riportate nella tabella 3.1.

Variabile	Stima	Errore Std.	z value	Pr(> z)
Intercept	-85.980	53.294	-1.613	0.107
sessoM	-1.343	0.604	-2.225	0.026*
altezza	0.960	0.604	1.590	0.112
altezza ²	-0.0028	0.0017	-1.615	0.106
peso	0.018	0.026	0.711	0.477
età	0.111	0.097	1.143	0.253

Table 3.1: Stime per il modello 3.1.

Nel modello attuale non si riscontra un effetto significativo dell'età sul dolore cervicale. Il risultato ha pienamente senso: la variabile età va dai 18 ai 25 anni, non ci aspettiamo una differenza significativa. La variabile Sesso appare significativa, anche se non è direttamente di interesse. Un uomo ha un log-rapporto delle quote stimato inferiore di -1.343 rispetto a una donna, a parità degli altri fattori. Quindi un uomo ha meno probabilità di soffrire di dolore al collo rispetto ad una donna.

3.2 Frequenza del dolore

Per le varie domande di ricerca sono stati stimati dei modelli per logit cumulati per risposte qualitative ordinali per la variabile $Y = \text{frequenza_dolore}$ con le seguenti modalità: '0 volte' (modalità $j = 1$); 'Sporadicamente (1-2 volte al mese)' (modalità $j = 2$); 'Frequentemente (almeno una volta a settimana)' (modalità $j = 3$); 'Quasi tutti i giorni'.

3.2.1 Frequenza del dolore e età

Sono stati inclusi nel modello le variabili di controllo sesso, altezza(quadratica) e peso, oltre a età, oggetto dello studio. Il modello è:

$$\text{logit}(P(Y \leq j)) = \theta_j + \beta_1 \cdot \text{sessoM} + \beta_2 \cdot \text{altezza} + \beta_3 \cdot \text{peso} + \beta_4 \cdot \text{età}, \quad j = 1, 2, 3, \quad (3.2)$$

dove θ_j sono le soglie (intercette).

Le stime vengono riportate nella tabella 3.2.

Variabile	Stima	Errore Std.	z value	Pr(> z)
(Intercept):1	4.097	3.548	1.155	0.248
(Intercept):2	5.880	3.568	1.648	0.099 .
(Intercept):3	7.048	3.584	1.966	0.049 *
sessoM	1.123	0.470	2.389	0.017 *
altezza	-0.0129	0.018	-0.707	0.480
peso	0.0056	0.021	0.274	0.784
età	-0.147	0.088	-1.663	0.096 .

Table 3.2: Stime per il modello 3.2.

Il parametro associato alla variabile età non risulta significativamente diverso da 0. Coerentemente con quanto stimato per il modello 3.1, per gli uomini si registra una maggiore probabilità di provare dolore con minore frequenza rispetto alle donne; ciò accade anche nei modelli successivi.

3.2.2 Frequenza del dolore e studio/pc

Il logit cumulato viene modellato con le variabili di controllo sesso, altezza, peso e età, oltre che alle ore passate a studiare o al pc.

Il modello è:

$$\text{logit}(P(Y \leq j)) = \theta_j + \beta_1 \cdot \text{sessoM} + \beta_2 \cdot \text{altezza} + \beta_3 \cdot \text{peso} + \beta_4 \cdot \text{età} + \beta_5 \cdot \text{ore_studio_e_pc}, \quad j = 1, 2, 3 \quad (3.3)$$

Le stime vengono riportate nella tabella 3.3.

Non è significativo l'effetto delle ore passate a studiare o al pc.

3.2.3 Frequenza del dolore e attività sportiva

Il logit cumulato viene modellato con le variabili di controllo sesso, altezza, peso e età, oltre che al numero di volte a settimana in cui la persona fa attività fisica.

Variabile	Stima	Errore Std.	z value	Pr(> z)
(Intercept):1	3.966	3.540	1.121	0.263
(Intercept):2	5.763	3.560	1.619	0.106 .
(Intercept):3	6.937	3.575	1.940	0.0523 .
sexoM	1.104	0.472	2.338	0.019 *
altezza	-0.014	0.018	-0.782	0.434
peso	0.007	0.021	0.340	0.734
età	-0.118	0.092	-1.283	0.199
ore_studio_e_pc	-0.092	0.081	-1.134	0.257

Table 3.3: Stime per il modello 3.3

Il modello è:

$$\text{logit}(P(Y \leq j)) = \theta_j + \beta_1 \cdot \text{sexoM} + \beta_2 \cdot \text{altezza} + \beta_3 \cdot \text{peso} + \beta_4 \cdot \text{età} + \beta_5 \cdot \text{freq_sport_quant}, \quad j = 1, 2, 3 \quad (3.4)$$

Le stime vengono riportate nella tabella 3.4.

Variabile	Stima	Errore Std.	z value	Pr(> z)
(Intercept):1	4.186	3.609	1.160	0.246
(Intercept):2	6.012	3.629	1.656	0.098 .
(Intercept):3	7.190	3.645	1.973	0.048 *
sexoM	1.101	0.473	2.325	0.020 *
altezza	-0.016	0.019	-0.879	0.380
peso	0.002	0.021	0.101	0.920
età	-0.133	0.090	-1.480	0.139
freq_sport_quant	0.219	0.107	2.040	0.041 *

Table 3.4: Stime per il modello 3.4

Il coefficiente relativo a **freq_sport_quant** risulta significativamente diverso da 0, tuttavia, siccome stiamo saggiando più ipotesi (come la nullità del coefficiente associato alle ore di studio del punto precedente), è opportuno utilizzare la correzione di Bonferroni per test multipli (o altre correzioni simili), per tenere sotto controllo la probabilità complessiva dell'errore del primo tipo. Con questa correzione, anche l'effetto dell'attività sportiva non risulta significativo.

3.2.4 Frequenza del dolore e utilizzo dello smartphone

Il logit cumulato viene modellato con le variabili di controllo sesso, altezza, peso e età, oltre che al numero di ore medie giornaliere di utilizzo dello smartphone o del tablet.

Il modello è:

$$\text{logit}(P(Y \leq j)) = \theta_j + \beta_1 \cdot \text{sexoM} + \beta_2 \cdot \text{altezza} + \beta_3 \cdot \text{peso} + \beta_4 \cdot \text{età} + \beta_5 \cdot \text{ore_smartphone_tablet}, \quad j = 1, 2, 3 \quad (3.5)$$

Le stime vengono riportate nella tabella 3.5.

Il coefficiente associato alle ore di utilizzo dello smartphone non è significativamente diverso da 0.

Variabile	Stima	Errore Std.	z value	Pr(> z)
(Intercept):1	3.670	3.645	1.007	0.314
(Intercept):2	5.464	3.663	1.492	0.136
(Intercept):3	6.635	3.678	1.804	0.071 .
sessom	1.191	0.478	2.492	0.013 *
altezza	-0.011	0.019	-0.585	0.559
peso	0.002	0.021	0.102	0.919
età	-0.153	0.088	-1.732	0.083 .
ore_smartphone_tablet	0.088	0.091	0.973	0.331

Table 3.5: Stime per il modello 3.5

3.2.5 Frequenza del dolore e altezza del dispositivo

Il logit cumulato viene modellato con le variabili di controllo sesso, altezza, peso e età, oltre che alla modalità di posizionamento dello smartphone (altezza del dispositivo), con valori possibili 'All'altezza del petto', 'All'altezza degli occhi o poco più in basso', 'All'altezza della pancia o più in basso'.

Il modello è:

$$\text{logit}(P(Y \leq j)) = \theta_j + \beta_1 \cdot \text{sessom} + \beta_2 \cdot \text{altezza} + \beta_3 \cdot \text{peso} + \beta_4 \cdot \text{età} + \beta_5 \cdot \mathbf{D}_{\text{petto}} + \beta_6 \cdot \mathbf{D}_{\text{pancia}}, \quad j = 1, 2, 3, \quad (3.6)$$

dove $D_{\text{petto}} = 1$ se il dispositivo è all'altezza del petto, 0 altrimenti; $D_{\text{pancia}} = 1$ se il dispositivo è all'altezza della pancia o più in basso, 0 altrimenti.

Le stime vengono riportate nella tabella 3.6.

Variabile	Stima	Errore Std.	z value	Pr(> z)
(Intercept):1	2.886	3.711	0.778	0.437
(Intercept):2	4.699	3.726	1.261	0.207
(Intercept):3	5.871	3.740	1.570	0.117
sessom	0.925	0.487	1.901	0.057 .
altezza	-0.005	0.019	-0.287	0.774
peso	0.007	0.021	0.333	0.739
età	-0.147	0.090	-1.635	0.102 .
altezza_dispositivo (petto)	-0.247	0.421	-0.586	0.558
altezza_dispositivo (pancia o più in basso)	0.806	0.653	1.233	0.218

Table 3.6: Stime per il modello 3.6

Il confronto di devianza (test del rapporto di verosimiglianza) con il modello ridotto (senza la covariata **altezza_dispositivo**) non mostra un miglioramento significativo: $\chi^2(2) = 3.2743$, $p = 0.1945$. Pertanto, non viene rifiutata l'ipotesi di nullità di entrambi i coefficienti associati all'altezza a cui è tenuto il dispositivo; ovvero non esiste una differenza significativa tra l'effetto di tenere il dispositivo all'altezza degli occhi (categoria di riferimento) e tenerlo in un'altra posizione.

3.3 Intensità del dolore

La domanda 'Solitamente, a che intensità può arrivare il dolore: (Indica un valore compreso tra 0 e 10, dove 0 equivale a nessun dolore e 10 equivale al massimo dolore)' prevedeva inizialmente una risposta numerica da 0 a 10. Per semplicità è stata ricodificata in 4 categorie, come illustrato nella tabella 3.7, e considerata come una variabile qualitativa ordinata. In maniera analoga a quanto fatto per la frequenza del dolore sono stati stimati dei modelli per logit cumulati.

Valore intensita_dolore	Classe assegnata (Y)
0	no_dolore
1-3	dolore_basso
4-6	dolore_medio
≥ 7	dolore_alto

Table 3.7: Schema di ricodifica di `intensita_dolore` in classi

3.3.1 Intensità e età

Il logit cumulato viene modellato con le variabili di controllo sesso, altezza e peso, oltre che alla variabile età. Il modello è:

$$\text{logit}(P(Y \leq j)) = \theta_j + \beta_1 \cdot \text{sessoM} + \beta_2 \cdot \text{altezza} + \beta_3 \cdot \text{peso} + \beta_4 \cdot \text{età}, \quad j = 1, 2, 3 \quad (3.7)$$

Le stime vengono riportate nella tabella 3.8.

Variabile	Stima	Errore Std.	z value	Pr(> z)
(Intercept):1	2.836	3.341	0.849	0.396
(Intercept):2	3.642	3.347	1.088	0.277
(Intercept):3	5.183	3.362	1.541	0.123
sessoM	1.416	0.453	3.122	0.002 **
altezza	-0.006	0.017	-0.341	0.733
peso	-0.015	0.019	-0.744	0.457
età	-0.093	0.084	-1.106	0.269

Table 3.8: Stime per il modello 3.7 sull'intensità del dolore cervicale.

L'effetto dell'età non risulta statisticamente significativo. Quello del sesso ancora una volta sì, sebbene non di interesse.

3.3.2 Intensità del dolore e ore di utilizzo dello smartphone

Il logit cumulato viene modellato con le variabili di controllo sesso, altezza, peso e età, oltre che al numero di ore medie giornaliere di utilizzo dello smartphone o del tablet.

Il modello è:

$$\text{logit}(P(Y \leq j)) = \theta_j + \beta_1 \cdot \text{sessoM} + \beta_2 \cdot \text{altezza} + \beta_3 \cdot \text{peso} + \beta_4 \cdot \text{età} + \beta_5 \cdot \text{ore_smartphone_tablet}, \quad j = 1, 2, 3 \quad (3.8)$$

Le stime vengono riportate nella tabella 3.9.

Variabile	Stima	Errore Std.	z value	Pr(> z)
(Intercept):1	2.673	3.398	0.787	0.431
(Intercept):2	3.479	3.403	1.022	0.307
(Intercept):3	5.020	3.418	1.469	0.142
sexoM	1.436	0.458	3.134	0.002 **
altezza	-0.005	0.017	-0.299	0.765
peso	-0.016	0.020	-0.796	0.426
età	-0.094	0.084	-1.123	0.261
ore_smartphone_tablet	0.034	0.088	0.383	0.702

Table 3.9: Stime per il modello 3.8 sull'intensità del dolore cervicale.

Le ore di utilizzo del dispositivo non hanno un effetto significativo sull'intensità del dolore provato.

3.3.3 Intensità del dolore e altezza del dispositivo

Il logit cumulato viene modellato con le variabili di controllo sesso, altezza, peso e età, oltre che alla modalità di posizionamento dello smartphone (altezza del dispositivo), con valori possibili "All'altezza del petto", "All'altezza degli occhi o poco più in basso", "All'altezza della pancia o più in basso".

Il modello è:

$$\text{logit}(P(Y \leq j)) = \theta_j + \beta_1 \cdot \text{sexoM} + \beta_2 \cdot \text{altezza} + \beta_3 \cdot \text{peso} + \beta_4 \cdot \text{età} + \beta_5 \cdot \mathbf{D}_{\text{petto}} + \beta_6 \cdot \mathbf{D}_{\text{pancia}}, \quad j = 1, 2, 3, \quad (3.9)$$

dove $D_{\text{petto}} = 1$ se il dispositivo è all'altezza del petto, 0 altrimenti; $D_{\text{pancia}} = 1$ se il dispositivo è all'altezza della pancia o più in basso, 0 altrimenti.

Le stime vengono riportate nella tabella 3.9.

Variabile	Stima	Errore Std.	z value	Pr(> z)
(Intercept):1	1.688	3.473	0.486	0.627
(Intercept):2	2.506	3.476	0.721	0.471
(Intercept):3	4.049	3.488	1.161	0.246
sexoM	1.260	0.467	2.698	0.007 **
altezza	-0.0001	0.017	-0.007	0.994
peso	-0.012	0.020	-0.592	0.554
età	-0.083	0.085	-0.973	0.331
altezza_dispositivo (petto)	-0.306	0.409	-0.750	0.453
altezza_dispositivo (pancia o più in basso)	0.497	0.626	0.794	0.427

Table 3.10: Stime per il modello 3.9 sull'intensità del dolore cervicale.

Il confronto di devianza (test del rapporto di verosimiglianza) con il modello ridotto (senza la covariata **altezza_dispositivo**) non mostra un miglioramento significativo: $\chi^2(2) = 2.0557$, $p = 0.357769$. Pertanto, non viene rifiutata l'ipotesi di nullità di entrambi i coefficienti associati all'altezza a cui è tenuto il dispositivo; ovvero non esiste una differenza significativa tra l'effetto di tenere il dispositivo all'altezza degli occhi (categoria di riferimento) e tenerlo in un'altra posizione per quanto riguarda l'intensità del dolore provato.

3.3.4 Intensità del dolore e ore di studio/pc

Il logit cumulato viene modellato con le variabili di controllo sesso, altezza, peso e età, oltre che alle ore passate a studiare o al pc.

Il modello è:

$$\text{logit}(P(Y \leq j)) = \theta_j + \beta_1 \cdot \text{sessoM} + \beta_2 \cdot \text{altezza} + \beta_3 \cdot \text{peso} + \beta_4 \cdot \text{età} + \beta_5 \cdot \text{ore_studio_e_pc}, \quad j = 1, 2, 3 \quad (3.10)$$

Le stime vengono riportate nella tabella 3.11.

Variabile	Stima	Errore Std.	z value	Pr(> z)
(Intercept):1	2.811	3.340	0.842	0.400
(Intercept):2	3.617	3.346	1.081	0.280
(Intercept):3	5.156	3.361	1.534	0.125
sessoM	1.400	0.455	3.079	0.002 **
altezza	-0.006	0.017	-0.368	0.713
peso	-0.014	0.020	-0.710	0.478
età	-0.084	0.088	-0.950	0.342
ore_studio_e_pc	-0.034	0.079	-0.431	0.667

Table 3.11: Stime per il modello 3.10 sull'intensità del dolore cervicale.

L'effetto delle ore passate a studiare o al pc non è significativo.

Bibliography

Wooldridge J. M. (2016). *Introductory econometrics a modern approach*. South-Western cengage learning.