

Tabagismo: analisi dei fattori di rischio

INDAGINE SUI FATTORI DI RISCHIO E DI PROTEZIONE APPARTENENTI AD AMBITI DEMOGRAFICI E MACROECONOMICI INFLUENZANTI LA PROBABILITÀ DI SVILUPPARE UNA DIPENDENZA DA FUMO DI SIGARETTA IN SOGGETTI CARATTERIALMENTE FORMATI

Domanda di ricerca e motivazione

Domanda di ricerca: “Indagare l’effetto di caratteristiche demografiche e macroeconomiche sulla probabilità di sviluppare una dipendenza da fumo da parte di soggetti mai stati fumatori di età compresa tra i 26 e 46 anni”

La **motivazione** sorge dall’incremento costante degli aspetti della vita quotidiana che generano una possibile dipendenza, dai bisogni primari come sostentamento, legami affettivi, possesso alle occasioni e modalità di svago, connessioni interpersonali e lavoro.

Si è scelto il range di età per evitare che le cause fossero confuse con il ***peer pressure effect***, tipico nei soggetti giovani.

Modalità di analisi

Metodologia: Data la natura della domanda di ricerca la metodologia scelta per l'analisi è quella della **survival analysis**, utilizzando modelli non parametrici e semi parametrici, prendendo in considerazione come tempo di sopravvivenza il periodo che intercorre tra l'inizio dello studio dove tutti i partecipanti hanno 26 anni e non sono mai stati fumatori ed un eventuale inizio di dipendenza da fumo.

Dataset

Il dataset finale è stato generato unendo vari follow-up compresi nel **British Cohort Study 1970**, una survey con cadenza quinquennale tutt'ora in corso su 17000 cittadini inglesi, disponibile previa concessione attraverso:

<https://census.ukdataservice.ac.uk/>

Il campione analizzato è rappresentativo di 8 etnie e controllato per includere tutto lo spettro di estrazioni sociali, condizioni economiche, caratteristiche demografiche e per assicurare che ogni soggetto non abbia mai fumato prima dei 26 anni, per un totale di 1743 soggetti.

Variabili

Variabili fisse nel tempo

Nome della variabile	dummy	significato	scala
female	si	1 = soggetto femmina	
etnia	no	1 = UK 2 = Ireland 3 = EU 4 = west indian 5 = other indian 6 = asian 7 = UK + other 8 = other	
Big 5	no	neuroticism	0 – 24
	no	openness	0 – 24
	no	extraversion	0 – 24
	no	conscientiousness	0 – 24
	no	agreeableness	0 – 24
smoking_hh	si	1 = esposizione ad un membro della famiglia fumatore	
draining	si	1 = sofferenza fisica o psicologica nel badare ai figli	
dummy_unemp	si	1 = ha subito periodi di lunga disoccupazione	
dummy_child_death	si	1 = ha subito la perdita di almeno un figlio	
dummy_div	si	1 = ha subito almeno un divorzio	
dummy_figli	si	1 = ha avuto figli	

Variabili

Variabili tempo dipendenti demografiche

Nome della variabile	dummy	significato	scala
age		età	26-46
empl	si	1 = periodo di disoccupazione negli ultimi 5 anni	
div	si	1 = almeno un divorzio negli ultimi 5 anni	
child_death	si	1 = morte di almeno un figlio negli ultimi 5 anni	

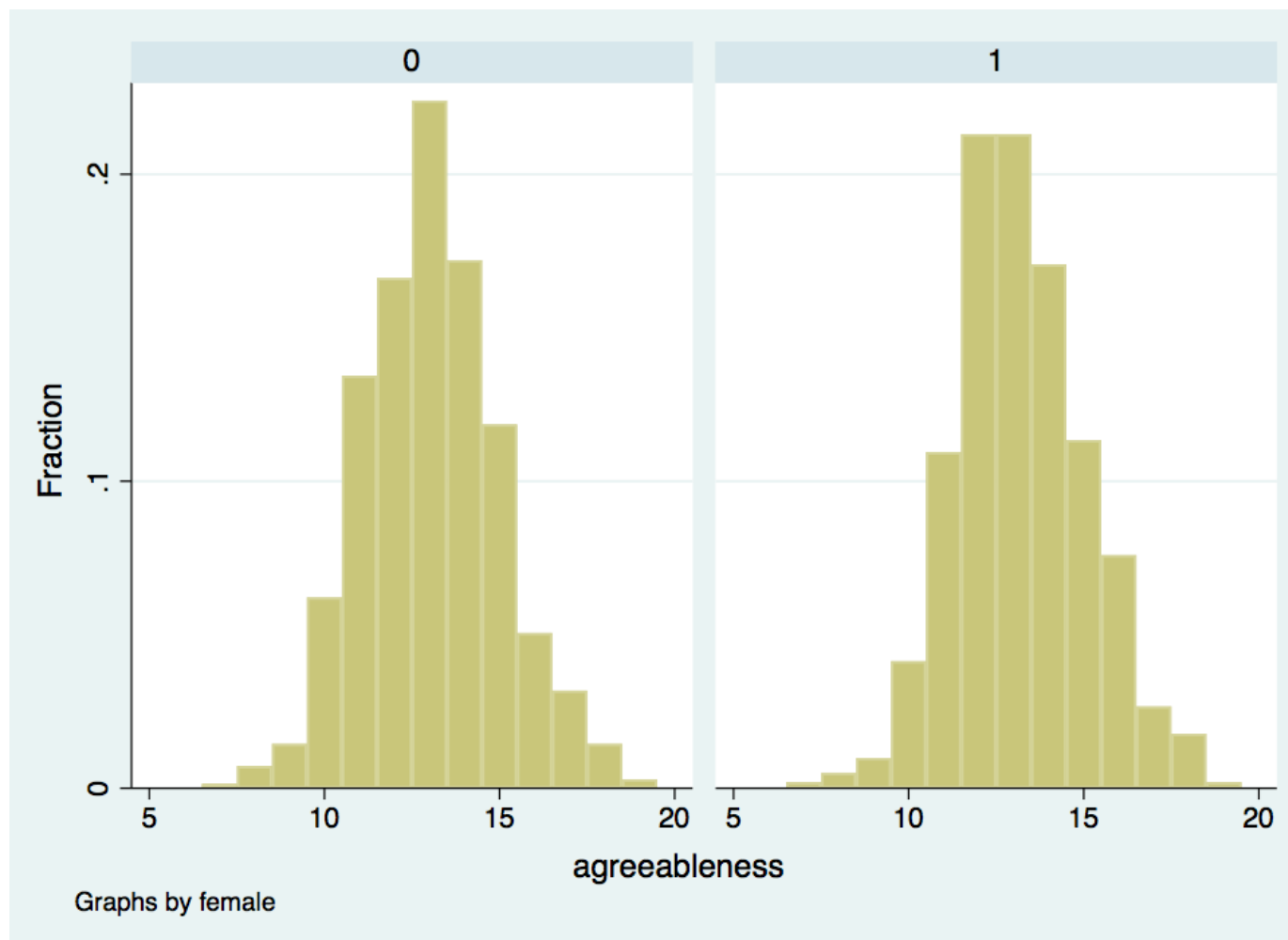
Variabili tempo dipendenti macroeconomiche – ogni 5 anni

Nome della variabile	significato
unempRate	tasso di disoccupazione
GDP	prodotto interno lordo
HhMRate	tasso di interesse sui mutui per la casa

Big 5

La **costruzione** delle cinque variabili è avvenuta vagliando manualmente tutte le centinaia di domande poste dai questionari in ogni follow up, per selezionare quelle che meglio corrispondono ai criteri previsti da (Christopher J. Soto, Oliver P. John 2017). La rappresentatività delle variabili generate è valutata positiva attraverso la coerenza delle 5 variabili con determinate caratteristiche demografiche, es:
Agreeableness femminile > agreeableness maschile

Big 5 agreeableness by gender



Come si può vedere, oltre alla tendenza alla normalità che in questo caso è particolarmente desiderabile per la conferma dell'adeguatezza delle variabili, è presente l'asimmetria tipica della distribuzione della variabile agreeableness tra maschi e femmine.

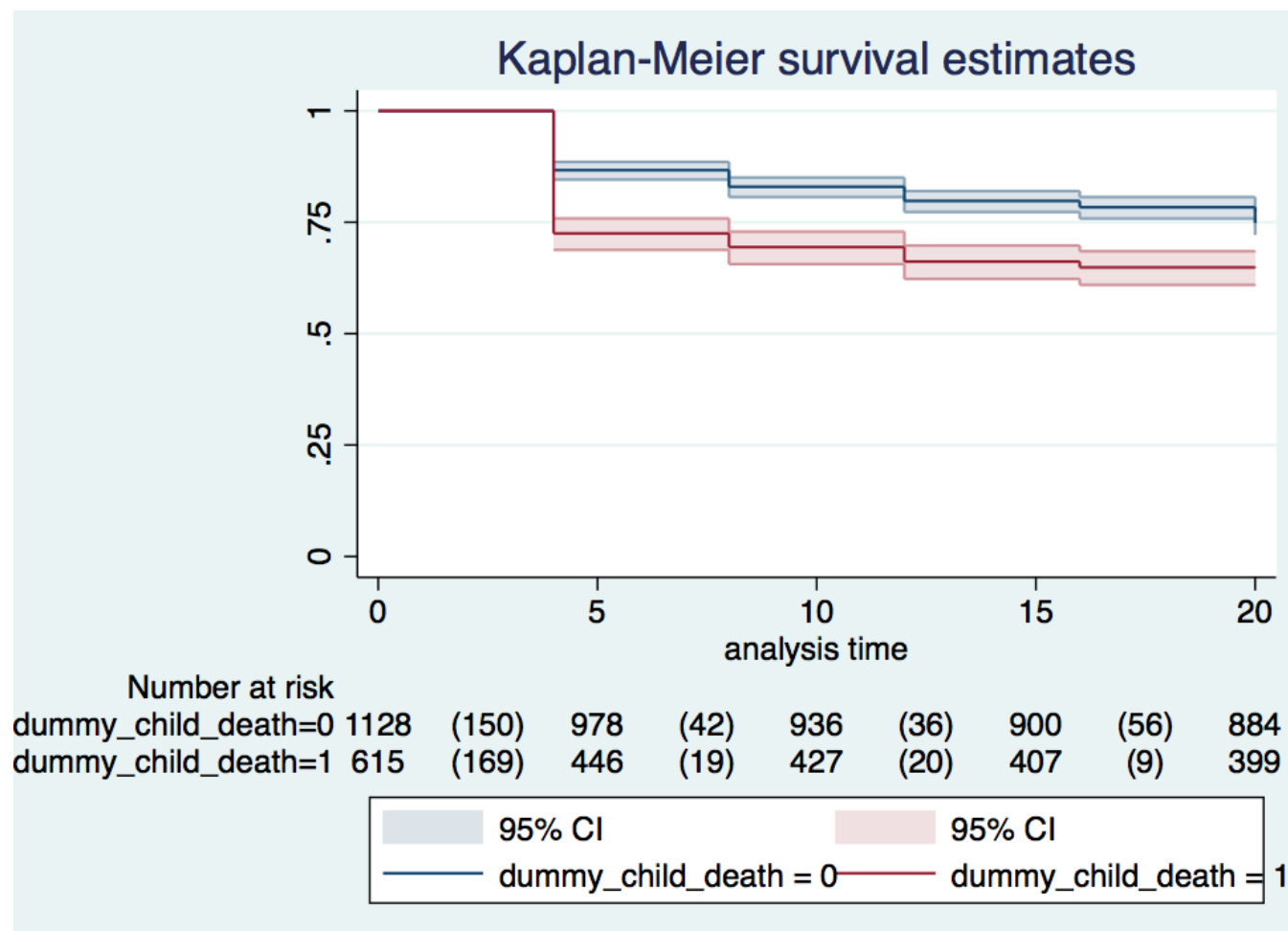
Log-rank test per le variabili dummy

Log-rank test variabili	H0: equality of survivor function P – value
female	0,020
smoking_hh	0,000
dummy_div	0,400
dummy_unempl	0,340
dummy_figli	0,120
draining	0,320
dummy_child_death	0,000

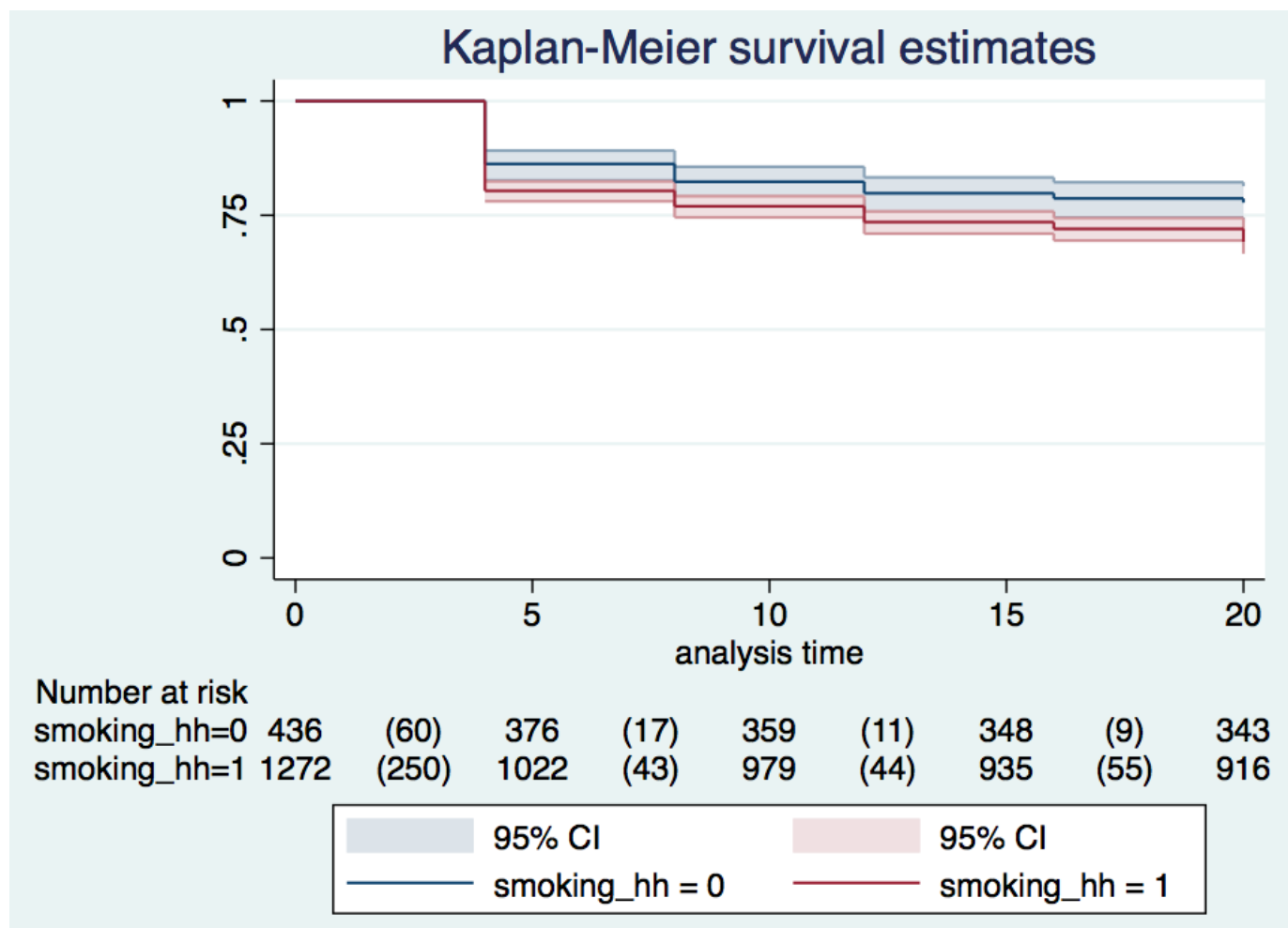
I risultati del log-rank test sono confermati dalle stime della funzione di sopravvivenza riportate in seguito

Risultati - Kaplan Meier – child death

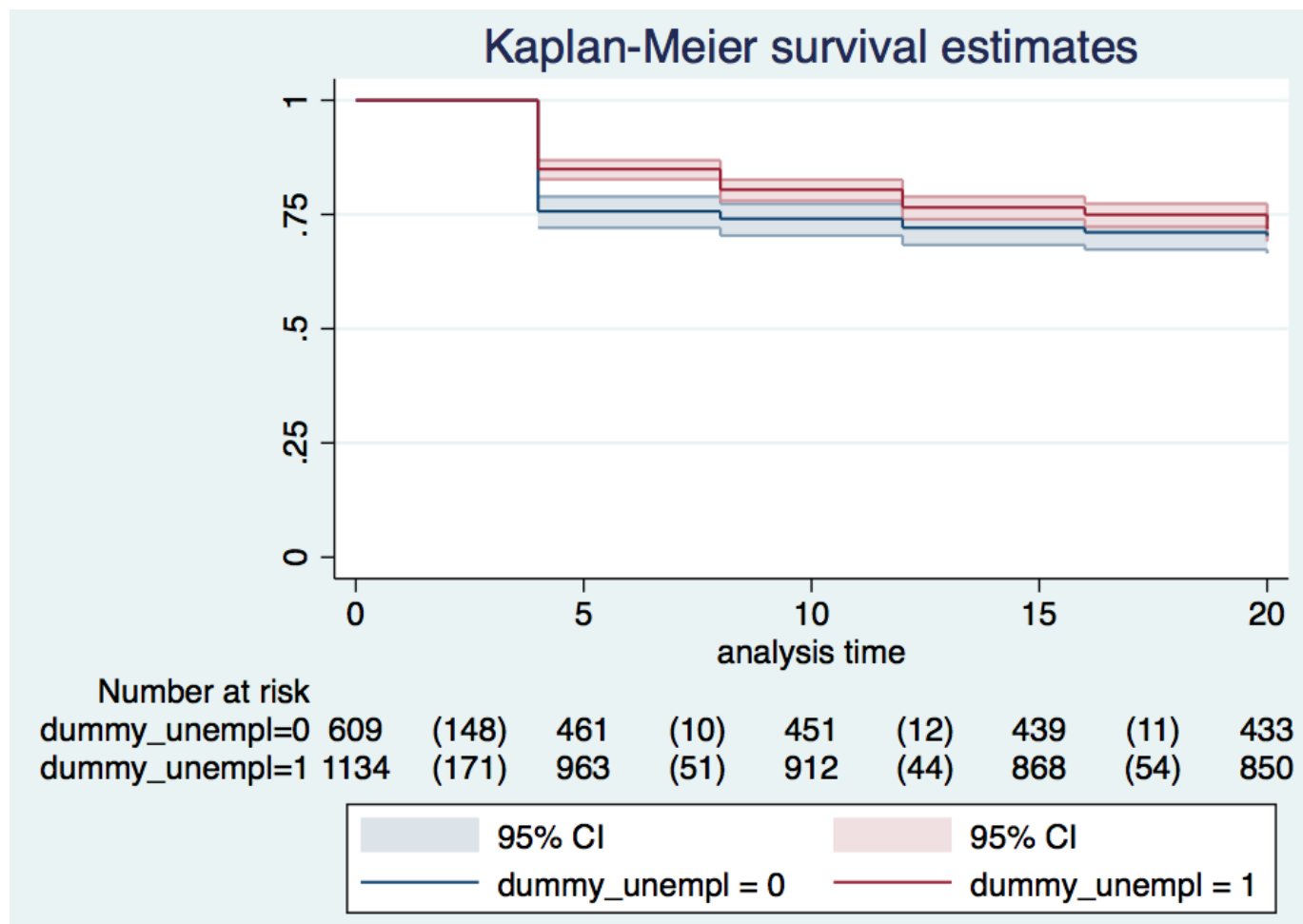
Di seguito sono riportate le stime della funzione di sopravvivenza per alcune stratificazioni significative



Risultati - Kaplan Meier – somking & relatives



Risultati - Kaplan Meier – ever unemployed



Risultati Cox model - main

_t	Haz. Ratio	Std. Err.	z	P> z
smoking_hh	1.37433	.1593374	2.74	0.006
extraversion	.9314268	.0194996	-3.39	0.001
conscientiousness	.9516038	.0225172	-2.10	0.036
openness	1.020929	.020386	1.04	0.300
neuroticism	1.034599	.030883	1.14	0.255
agreeableness	.9620298	.0236292	-1.58	0.115
dummy_child_death	1.610457	.1536736	4.99	0.000
female	.8538858	.0824937	-1.64	0.102
empl	.5957674	.1553212	-1.99	0.047

E' stato controllato che: il modello sia stato specificato correttamente, non vi siano punti ad alta leva, sia rispettata l'assunzione di proporzionalità, sia accettabile la bontà di adattamento.

tvc		P-values				
age	.9996317
unempRate	1	205501.7	-0.00	1.000	0	.
GDP	1	.5543199	0.00	1.000	.3374132	2.963725
HhMRate	1	(omitted)				

Note: Variables in tvc equation interacted with (_t).

Vi sono problemi per alcune variabili tempo dipendenti

Risultati Cox model - Stratification

Strat by	Variable	Hazard Rateo	P-value
Female = 0	dummy_child_death	1,760	0,000
Female = 1	dummy_child_death	1,480	0,002
smoking_hh = 0	extraversion	0,820	0,000
smoking_hh = 1	extraversion	0,950	0,059
dummy_unemp = 0	smoking_hh	1,410	0,084
dummy_unemp = 1	smoking_hh	1,360	0,029
draining = 0	dummy_child_death	1,550	0,000
draining = 1	dummy_child_death	1,750	0,002
dummy_child_death = 0	smoking_hh	1,290	0,074
dummy_child_death = 1	smoking_hh	1,480	0,043

Qui sono evidenziate delle differenze significative negli hazard rateo rispetto alla variabile della colonna **Variable** quando si stratifica per la variabile definita nella colonna **Strat by**

Conclusioni

L'aver perso un figlio e essere cresciuti in un ambiente dove il fumo era consueto si sono rivelati i due più forti fattori di rischio.

Per ciò che concerne i tratti caratteriali, essere estroverso e accomodante, così come essere diligente sono tutti fattori protettivi rispetto all'insorgenza del tabagismo.

Le donne e le persone con un'occupazione stabile corrono un rischio minore di sviluppare la patologia in esame

I pochi intervalli temporali a disposizione sono la causa più verosimile dell'impossibilità di includere nel modello le variabili tempo dipendenti continue

Conclusioni

Grazie per l'attenzione e per la collaborazione durante questo progetto.

