

Analisi esplorativa e statistica del consumo globale di alcol: evidenze da dati WHO disaggregati per fattori socio-demografici

Un approccio basato su EDA e test statistici per l'analisi delle disuguaglianze nel consumo di alcol

Daniele Angeloni
Università degli Studi di Perugia
daniele.angeloni@studenti.unipg.it

Alessandra Ruggeri
Università degli Studi di Perugia
alessandra.ruggeri@studenti.unipg.it

Abstract—Il presente lavoro esplora le disuguaglianze globali nel consumo di alcol attraverso l'analisi di dati disaggregati provenienti dall'*Health Inequality Data Repository* della *World Health Organization (WHO)*. L'indagine si è focalizzata su tre indicatori chiave: la percentuale di consumatori negli ultimi 12 mesi, la percentuale di astinenti a vita e il consumo pro capite (in litri di alcol puro) tra i soli bevitori. I dati sono stati sottoposti a una fase di preprocessing e analisi in linguaggio R, con particolare attenzione alle differenze socio-demografiche per genere, fascia di reddito, regione geografica e andamento temporale negli anni.

L'analisi esplorativa ha messo in evidenza pattern significativi: una prevalenza marcata del consumo tra gli uomini, una correlazione positiva tra livello di reddito e intensità del consumo e forti disomogeneità tra le diverse regioni OMS. Per valutare la significatività delle differenze osservate sono stati applicati test statistici non parametrici (Wilcoxon, Kruskal-Wallis, Friedman), preceduti da test di normalità e accompagnati da analisi post-hoc correttive dove necessario.

I risultati ottenuti confermano la rilevanza dei dati WHO per lo studio comparativo di comportamenti sanitari a livello globale e sottolineano la necessità di politiche sanitarie differenziate e mirate per contrastare le disuguaglianze legate a determinanti socio-demografici.

I. INTRODUZIONE

Il consumo di alcol rappresenta un importante indicatore di disuguaglianza nella salute pubblica globale, in quanto riflette differenze significative nei comportamenti e negli esiti sanitari tra gruppi socioeconomici, culturali e demografici.

Tali disuguaglianze si manifestano non solo nella diffusione del consumo, ma anche nella sua intensità e nelle conseguenze sulla salute individuale e collettiva. L'analisi si è articolata in due fasi principali: un'esplorazione descrittiva preliminare per identificare tendenze e disuguaglianze nei dati, seguita da un'analisi statistica inferenziale finalizzata a verificare la significatività delle differenze osservate tra gruppi socio-demografici. L'intero processo è stato condotto in linguaggio R, attraverso operazioni di preprocessing, esplorazione dei dati, visualizzazione e applicazione di test statistici.

II. DATASET E METODOLOGIA

Il dataset analizzato, estratto dall'*Health Inequality Data Repository* della WHO [1], contiene misurazioni sanitarie

disaggregate per età, sesso, gruppo di reddito e regione geografica, con una copertura temporale complessiva che va dal 1950 al 2023.

Ai fini di questa analisi, sono stati selezionati tre indicatori specifici relativi al consumo di alcol, per i quali i dati erano disponibili solo per il periodo 2000–2020 [2]:

- `Alcohol, consumers past 12 months (%)`: percentuale di persone che hanno consumato alcolici negli ultimi 12 mesi;
- `Alcohol, abstainers lifetime (%)`: percentuale di persone che non hanno mai consumato alcol nella vita;
- `Alcohol, drinkers only per capita (15+) consumption (litres of pure alcohol)`: litri medi di alcol puro consumati pro capite tra i soli bevitori.

Dopo l'importazione del dataset in formato XLSX, i dati sono stati preprocessati mediante:

- selezione delle variabili rilevanti e manipolazione dei nomi delle etichette per renderle più interpretabili (`country`, `year`, `gender`, `income_group`, `who_region`);
- gestione dei valori mancanti: le righe con valore mancante nella variabile `value` sono state rimosse, trattandosi di un numero esiguo; per la variabile `income_group`, i valori mancanti sono stati mantenuti e riclassificati come `Unknown`.

L'analisi si è articolata in due fasi principali. La prima, di tipo esplorativo, ha previsto la produzione di visualizzazioni descrittive per identificare pattern ricorrenti, variazioni temporali e possibili anomalie locali.

La seconda fase, di tipo inferenziale, ha incluso:

- applicazione del test di normalità di Shapiro-Wilk [3] per ciascuna combinazione variabile-indicatore;
- utilizzo di test statistici non parametrici (Wilcoxon [4], Kruskal-Wallis [5], Friedman [6]);
- analisi post-hoc con test di Dunn [7] e correzione di Bonferroni [8] (dove necessario).

III. ANALISI ESPLORATIVA DEI DATI

L'analisi esplorativa dei dati (EDA) ha lo scopo di investigare ed esplorare la struttura e le tendenze presenti nel dataset relativo al consumo di alcol.

L'analisi è stata articolata in più livelli: un controllo dei range di valori associati agli indicatori per verificarne la validità; una panoramica generale basata su media, deviazione standard e tendenze globali; un approfondimento per fascia di reddito; una valutazione delle variazioni temporali a livello regionale; l'identificazione di possibili anomalie nei comportamenti di consumo; e, infine, uno studio del gender gap nel consumo. Ogni sottosezione è corredata da visualizzazioni grafiche mirate, volte a facilitare l'interpretazione dei fenomeni osservati.

A. Analisi generale

L'analisi esplorativa si apre con una panoramica sui tre indicatori considerati: Litres_Cons, Perc_Ast e Perc_Cons, disaggregati per genere.

Le medie e deviazioni standard calcolate separatamente per uomini e donne mostrano una netta differenza tra i sessi: gli uomini presentano un consumo medio di 16.63 litri contro i 6.11 litri delle donne. Anche la percentuale di consumatori è maggiore tra gli uomini, mentre quella di astinenti è più alta tra le donne. Le differenze sono state visualizzate tramite un barplot.

L'analisi delle correlazioni (Spearman [9]) ha evidenziato una forte relazione negativa tra la percentuale di astinenti e quella di consumatori. Tali risultati sono coerenti con il fatto che nei contesti in cui l'astinenza è diffusa, la quota di consumatori è inevitabilmente ridotta.

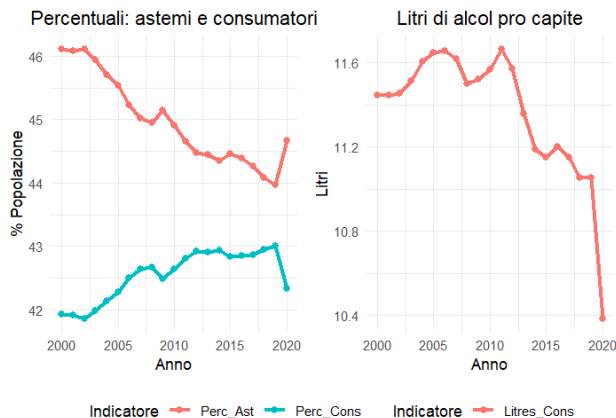


Fig. 1. Trend globale nel tempo per indicatore

Per comprendere le dinamiche temporali aggregate, sono stati tracciati gli andamenti annuali dei tre indicatori, mediati per genere, tramite *line chart*. La percentuale di astinenti è globalmente in diminuzione mentre quella di consumatori è globalmente in crescita, mentre il consumo medio tra i bevitori mostra un trend decrescente a partire dal 2010, suggerendo una riduzione dell'intensità del consumo. Da notare che nel 2020

si visualizza una notevole inversione della tendenza globale, probabilmente dovuto all'epidemia di Covid-19. (Fig. 1).

Infine, la distribuzione dei tre indicatori per regione geografica è stata rappresentata tramite dei *violin plot*, evidenziando forti disomogeneità: la regione EURO mostra i valori più elevati in termini di consumo pro capite tra i bevitori, mentre EMRO si distingue per l'elevata percentuale di astinenti. AMRO ed EURO presentano le quote più alte di consumatori regolari, a differenza di EMRO e SEARO, dove prevale l'astinenza. Si può presupporre che queste disuguaglianze siano anche dovute a motivazioni culturali e/o di orientamento religioso della zona geografica interessata.

B. Analisi per fascia di reddito (Income Group)

L'analisi è stata estesa considerando la suddivisione dei paesi secondo le fasce di reddito definite dalla Banca Mondiale: *Low*, *Lower-middle*, *Upper-middle* e *High income*. Per ciascun gruppo sono stati calcolati media e deviazione standard dei tre indicatori principali.

I risultati mostrano un pattern netto: nei paesi ad alto reddito si rilevano in media valori più elevati sia nella *percentuale di consumatori* sia nei *litri di alcol pro capite* tra i bevitori, mentre nei paesi a basso reddito è molto più diffusa l'*astinenza totale*. Nei gruppi intermedi (*lower* e *upper middle income*) si osserva una maggiore variabilità interna, soprattutto nella percentuale di astinenti.

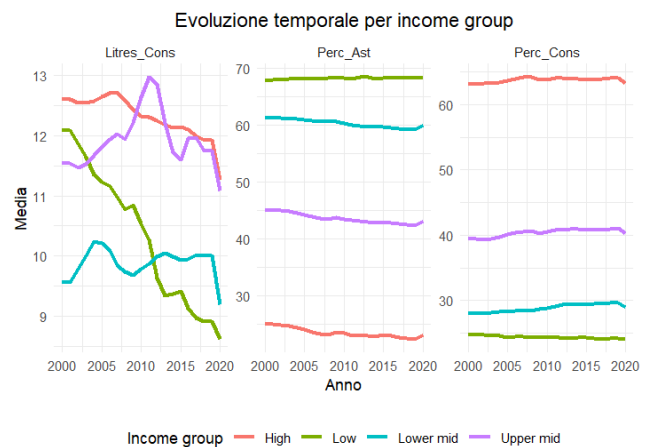


Fig. 2. Evoluzione temporale dei valori medi per income group

L'evoluzione temporale degli indicatori, rappresentata tramite *line chart*, mostra che le percentuali di astinenti a vita e di consumatori, restano complessivamente stabili nel tempo, o comunque presentano tendenze lievi di crescita o decrescita, per tutte le fasce di reddito. Al contrario, il consumo medio pro capite tra i bevitori evidenzia una tendenza alla diminuzione, con un calo particolarmente marcato nei paesi a basso reddito (Fig. 2).

Tali evidenze confermano l'influenza del contesto socio-economico sulle abitudini alcoliche e motivano l'uso di test statistici per valutarne la significatività.

C. Variazioni annuali regionali

È stata condotta un'analisi della variazione percentuale anno su anno della *percentuale di consumatori* (Fig. 3) e della *percentuale di astinenti* a livello regionale (WHO), con l'obiettivo di mettere in evidenza fluttuazioni locali che non emergono nei trend globali aggregati.

Le variazioni sono state calcolate come differenze percentuali rispetto all'anno precedente, per ciascuna combinazione anno-regione. Le analisi, rappresentate originariamente tramite *heatmap*, hanno permesso di individuare variazioni ricorrenti e anomalie specifiche.

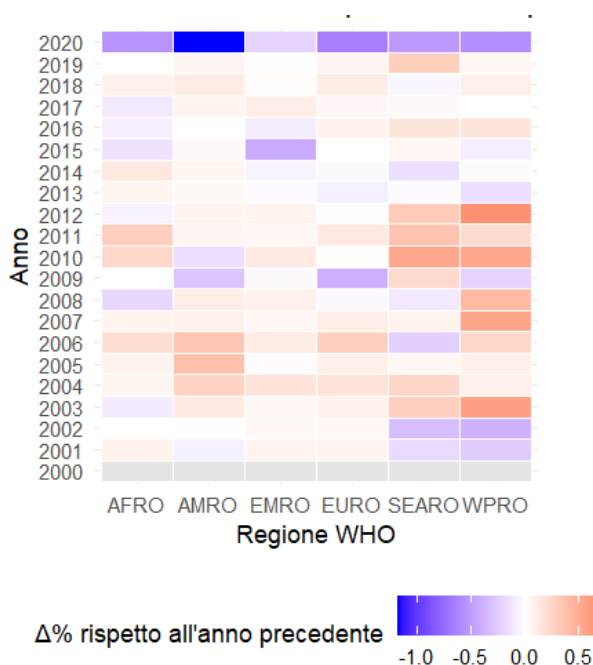


Fig. 3. Variazioni regionali della percentuale di consumatori rispetto all'anno precedente

I risultati evidenziano pattern regionali ben distinti. Nel 2020, ad esempio, la regione AFRO ha registrato un calo marcato nella percentuale di consumatori, accompagnato da un aumento netto della quota di astinenti, segnalando una potenziale discontinuità comportamentale o normativa. La regione WPRO ha mostrato un'aumento progressiva dei consumatori tra il 2003 al 2012, seguita da alcune controtendenze locali. Al contrario, la regione SEARO si è caratterizzata per fluttuazioni meno sistematiche e più irregolari.

Si può comunque notare come nel 2020 in tutte le regioni ci sia stato un calo netto della percentuale dei consumatori rispetto all'anno precedente.

Queste evidenze suggeriscono che, accanto alle tendenze globali, esistono dinamiche regionali autonome, potenzialmente legate a fattori culturali, sanitari, a interventi di policy pubblica o a eventi eccezionali come la pandemia da COVID-19 nel 2020, che potrebbe aver influenzato i comportamenti di consumo.

D. Anomalie e paradosso del consumo

Un'analisi mirata è stata condotta per identificare comportamenti anomali nel rapporto tra *astinenza a vita* e *consumo medio tra i bevitori*. A tal fine è stato realizzato uno *scatterplot* (Fig. 4) che mette in relazione la percentuale di astinenti con i litri medi di alcol puro consumati tra chi beve, rappresentando i Paesi secondo la loro regione WHO di appartenenza.

Dal grafico emergono chiaramente alcuni outlier. In particolare, un Paese della regione EMRO mostra una combinazione estrema: oltre il 90% di astinenti e più di 30 litri pro capite tra i bevitori. Questo fenomeno rivela una struttura polarizzata della popolazione, con una piccola minoranza che consuma quantità elevate di alcol e una maggioranza che si astiene completamente.

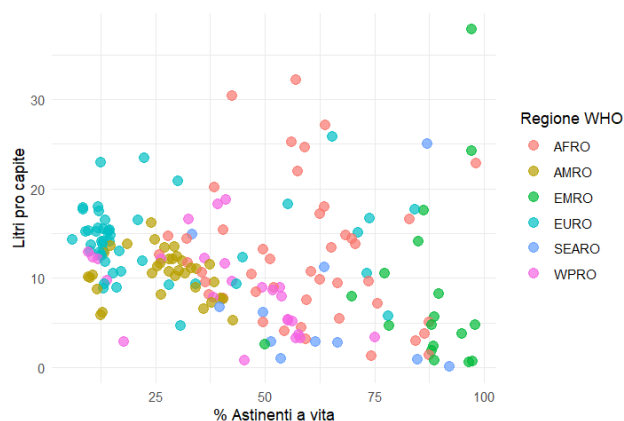


Fig. 4. Astinenti (%) vs Litri pro capite (tra i bevitori)

Per quantificare questa dinamica è stato introdotto un *indice di paradosso*, definito come il rapporto tra il consumo medio tra i bevitori e la percentuale di astinenti. L'indice ha permesso di identificare i Paesi con la coesistenza più marcata tra astinenza diffusa e consumo elevato. Tra i primi dieci rientrano principalmente Stati a prevalenza islamica (es. Siria, Comore, Iran), confermando una localizzazione geografica nella regione EMRO.

E. Differenze di Genere (Gender Gap)

L'analisi ha considerato le differenze di genere nel consumo di alcol tra i soli bevitori, evidenziando lo stesso pattern: in tutte le regioni WHO, gli uomini consumano quantità mediamente superiori rispetto alle donne. Il divario è particolarmente marcato nella regione EURO, dove il consumo maschile raggiunge i livelli più elevati. Invece, nella regione SEARO, la differenza relativa è meno accentuata.

Per valutare l'evoluzione della disparità nel tempo, è stato calcolato il *gender gap* come differenza media annuale tra i litri consumati da uomini e donne, suddivisa per regione. I risultati, visualizzati tramite *line chart* (Fig. 5), mostrano una certa eterogeneità tra aree geografiche: in alcune regioni (come SEARO, AFRO e AMRO) il gap rimane stabile, mentre in altre (es. EMRO) si osserva una forte riduzione, suggerendo

un possibile riavvicinamento dei livelli di consumo. Infine in WPRO possiamo osservare un aumento del gap.

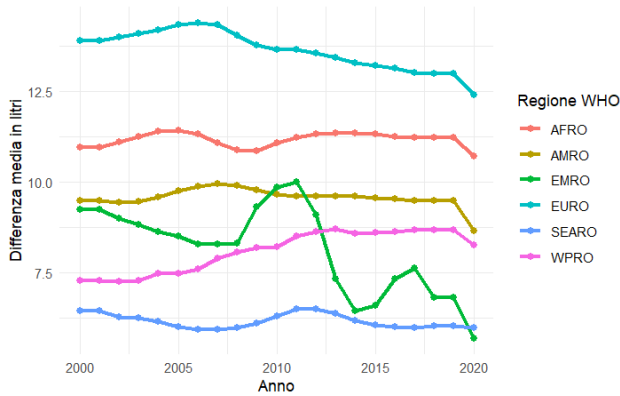


Fig. 5. Gender gap nel consumo di alcol tra i bevitori (litri pro capite)

Nel complesso, nonostante alcune variazioni locali, il divario tra uomini e donne persiste in modo strutturale su scala globale. Questo indica che il comportamento alcolico rimane profondamente differenziato per genere, probabilmente in relazione a fattori culturali, sociali, normativi e anche biologici, considerando che le donne tendono a consumare meno alcol sia per norme sociali sia per una diversa risposta fisiologica agli effetti dell'alcol.

IV. ANALISI STATISTICA

Dopo la fase esplorativa, è stata condotta un'analisi statistica inferenziale per verificare se le differenze osservate tra gruppi socio-demografici ed economici fossero statisticamente significative. L'analisi si è focalizzata sui tre indicatori principali rispetto alle variabili di raggruppamento.

A. Test di normalità

Per stabilire l'idoneità all'utilizzo di eventuali test parametrici, è stata verificata come prima assunzione necessaria la normalità delle distribuzioni per ciascun indicatore (*Litres_Cons*, *Perc_Cons*, *Perc_Ast*) rispetto alle principali variabili di raggruppamento: *gender*, *income_group* e *who_region*. A tal fine, è stato applicato il test di Shapiro-Wilk [3], considerato affidabile per campioni di piccola e media dimensione, unitamente a un'analisi visiva mediante QQ-plot.

I risultati hanno evidenziato in modo sistematico valori di $p < 0.05$ nella maggior parte dei confronti, indicando una significativa deviazione dalla normalità teorica. I QQ-plot confermano queste evidenze, mostrando distorsioni rispetto alla diagonale di riferimento, code pesanti e distribuzioni asimmetriche. Tali anomalie sono particolarmente evidenti nei dati aggregati per regioni WHO e fasce di reddito, dove la presenza di outlier e forti eterogeneità è più marcata.

Sulla base di questi risultati, è stata esclusa l'applicazione di test parametrici (come il t-test o l'ANOVA), optando per test non parametrici più robusti, che non richiedono assunzioni stringenti sulla distribuzione dei dati. La scelta è inoltre

coerente con la natura osservazionale e disomogenea del dataset WHO, che aggrega dati eterogenei provenienti da numerosi paesi, con dinamiche socio-economiche e culturali molto diverse.

Le tabelle dettagliate e i grafici QQ-plot relativi ai singoli test sono disponibili e consultabili nel materiale supplementare del progetto, di seguito un esempio:

TABLE I
RISULTATI TEST SHAPIRO-WILK PER *LITRES_CONS* RISPETTO A *GENDER*

gruppo	categoria	n	W	p_value
gender	Female	3835	0.894	1.500000e-45
gender	Male	3835	0.897	4.170000e-45

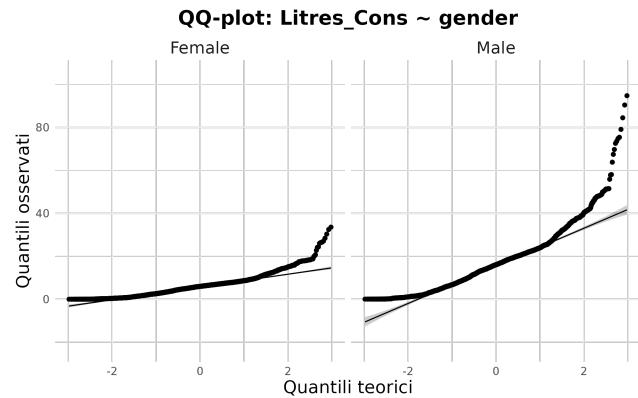


Fig. 6. QQ-plot per *Litres_Cons* rispetto a *gender*

B. Test Statistici Non Parametrici

A seguito delle verifiche di normalità tutte le combinazioni tra indicatori e variabili di raggruppamento hanno mostrato distribuzioni non normali ($p < 0.05$), motivando l'impiego di test statistici non parametrici. Di seguito vengono presentati i test effettuati:

Consumo medio tra bevitori (Litres_Cons) per genere: Per confrontare il consumo medio pro capite tra i soli bevitori (*Litres_Cons*) nei due generi, è stato applicato il **Wilcoxon rank-sum test** [4].

TABLE II
RISULTATI DEL TEST WILCOXON PER
WILCOXON_LITRES_CONS_BY_GENDER

Test	Statistica	df	p-value
Wilcoxon	2192242	-	< 2.2e-16

Il risultato è altamente significativo ($p < 0.001$), confermando che gli uomini consumano in media quantità significativamente maggiori di alcol rispetto alle donne. Questo divario è confermato anche dall'analisi esplorativa.

Percentuale di consumatori (Perc_Cons) per fascia di reddito: Per valutare se la prevalenza dei consumatori negli ultimi 12 mesi (*Perc_Cons*) varia in base alla fascia di

reddito, è stato utilizzato il **Kruskal-Wallis test** [5], seguito da post-hoc di Dunn con Correzione di Bonferroni [7] [8].

TABLE III
RISULTATI DEL TEST KRUSKAL PER
KRUSKAL_PERC_CONS_BY_INCOME_GROUP

Test	Statistica	df	p-value
Kruskal-Wallis	2815.8	3	< 2.2e-16

TABLE IV
RISULTATI DEL TEST DUNN PER
DUNN_PERC_CONS_BY_INCOME_GROUP

Comparison	Z	P_unadj	P_adj
High income - Low income	41.594869	0.00e+00	0.00e+00
High income - Lower middle income	45.685065	0.00e+00	0.00e+00
Low income - Lower middle income	-4.672610	2.97e-06	1.78e-05
High income - Upper middle income	30.658551	2.03e-206	1.22e-205
Low income - Upper middle income	-16.464539	6.60e-61	3.96e-60
Lower middle income - Upper middle income	-14.442105	2.81e-47	1.69e-46

I risultati mostrano differenze significative ($p < 0.001$), in particolare tra paesi ad alto reddito e quelli a reddito basso. Nei primi, la quota di consumatori risulta sistematicamente più alta. Il grafico seguente evidenzia chiaramente questa netta differenza:

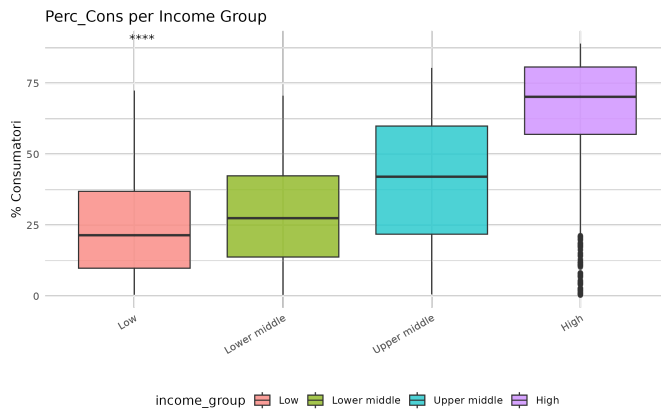


Fig. 7. Percentuale di consumatori (Perc_Cons) per fascia di reddito

Percentuale di astinenti a vita (Perc_Ast) per regione WHO: Un'ulteriore analisi ha riguardato la percentuale di astinenti a vita (Perc_Ast) nelle sei regioni WHO. Il **Kruskal-Wallis test** [5] ha indicato differenze altamente significative tra le regioni, come confermato anche dal test di Dunn [7].

TABLE V
RISULTATI DEL TEST KRUSKAL PER
KRUSKAL_PERC_AST_BY_WHO_REGION

Test	Statistica	df	p-value
Kruskal-Wallis	4280	5	< 2.2e-16

TABLE VI
RISULTATI DEL TEST DUNN PER PERC_AST RISPETTO A WHO_REGION

Comparison	Z	P_unadj	P_adj
AFRO - AMRO	33.500610	4.72e-246	7.08e-245
AFRO - EMRO	-20.516810	1.52e-93	2.29e-92
AMRO - EMRO	-46.798042	0.00e+00	0.00e+00
AFRO - EURO	42.163805	0.00e+00	0.00e+00
AMRO - EURO	5.159904	2.47e-07	3.71e-06
EMRO - EURO	53.904717	0.00e+00	0.00e+00
AFRO - SEARO	-1.588806	1.12e-01	1.00e+00
AMRO - SEARO	-23.281419	6.84e-120	1.03e-118
EMRO - SEARO	13.084699	4.03e-39	6.04e-38
EURO - SEARO	-27.542662	5.42e-167	8.13e-166
AFRO - WPRO	16.258830	1.93e-59	2.90e-58
AMRO - WPRO	-13.267709	3.56e-40	5.34e-39
EMRO - WPRO	31.902098	2.50e-223	3.75e-222
EURO - WPRO	-18.782016	1.06e-78	1.59e-77
SEARO - WPRO	12.639499	1.28e-36	1.92e-35

Le regioni EMRO e SEARO si distinguono per alti livelli di astinenza, a differenza di EURO e AMRO, dove prevale il consumo. La figura seguente illustra la distribuzione delle percentuali per area geografica:

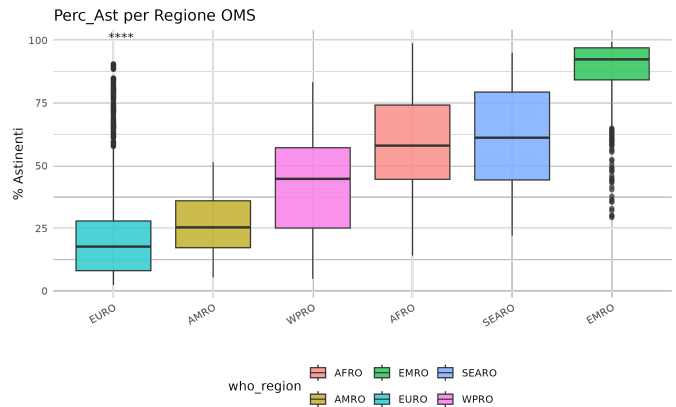


Fig. 8. Astinenti a vita (Perc_Ast) per regione WHO

Trend temporali del consumo tra paesi europei: Per analizzare l'evoluzione temporale dell'indicatore Litres_Cons in tre paesi europei ad alto reddito (Italia, Francia, Germania) è stato adottato il **Friedman test**, scelta motivata dal fatto che il dataset presenta, per ciascun paese (considerato come soggetto), 21 misurazioni annuali tra il 2000 e il 2020.

L'obiettivo dell'analisi è verificare se il consumo medio ha mostrato variazioni significative tra gli anni, tenendo conto che le osservazioni tra anni successivi sono dipendenti (misure ripetute) e che il numero di paesi considerato è limitato a tre, rendendo inappropriati approcci parametrici.

TABLE VII
RISULTATI DEL TEST FRIEDMAN PER
FRIEDMAN_LITRES_ITALY_FRANCE_GERMANY

Test	Statistica	df	p-value
Friedman	47.031	20	0.0005809

Il test ha restituito un valore di $p < 0.05$, indicando la presenza di almeno un anno in cui il consumo medio si discosta significativamente rispetto agli altri, considerando l'andamento dei tre paesi nel loro insieme.

Per esplorare più nel dettaglio in quali anni si concentrano le variazioni, è stato eseguito un **test post-hoc per Friedman**, con correzione di Bonferroni.

TABLE VIII
RISULTATI DEL TEST POST-HOC DI FRIEDMAN PER LITRES_CONS TRA ITALY, FRANCE E GERMANY (ANNI 2000–2020)

Group1	Group2	p-value
2015	2000	0.219728
2016	2000	0.153810
2020	2000	0.000040
2016	2002	0.763016
2020	2002	0.005238
2020	2003	0.028373
2020	2004	0.105238
2020	2005	0.307619

Nota: sono stati esclusi dalla tabella i confronti con p-value non calcolabile (NaN) e quelli con p-value = 1.

I risultati mostrano che il 2020 si distingue significativamente rispetto ad altri anni, alla luce di eventi come la pandemia di COVID-19, che ha influenzato comportamenti sociali e abitudini di consumo in tutta Europa (specialmente in paesi come Italia, Francia e Germania che hanno adottato misure di contenimento simili). Dal line-plot invece si nota un trend non omogeneo negli anni del consumo medio nei tre paesi presi come riferimento.

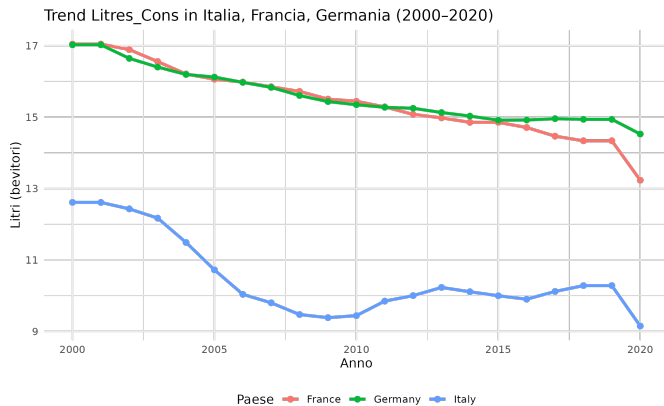


Fig. 9. Trend 2000–2020 del consumo medio (Litres_Cons) in Italia, Francia e Germania

Oltre ai quattro test presentati, sono stati condotti altri due test con metodologia analoga: un *Wilcoxon rank-sum test* [4] applicato alla percentuale di consumatori (Perc_Cons) tra uomini e donne ha evidenziato una differenza significativa ($p < 0.001$), confermando una maggiore prevalenza maschile. Analogamente, un *Kruskal-Wallis test* [5] sulla variabile Litres_Cons per fascia di reddito ha mostrato un incremento statisticamente significativo del consumo medio al

crescere della ricchezza nazionale. Le tabelle e i grafici relativi a queste analisi sono disponibili nel materiale supplementare del progetto.

V. CONCLUSIONI E SVILUPPI FUTURI

L'analisi ha messo in luce profonde disuguaglianze nel consumo di alcol su scala globale, legate a fattori socio-demografici ed economici come genere, reddito e appartenenza geografica. In tutte le regioni geografiche, gli uomini risultano consumatori più assidui e intensi rispetto alle donne; nei paesi ad alto reddito si rilevano livelli più elevati di consumo, mentre in quelli a basso reddito prevale l'astinenza. A livello regionale, si osservano marcate differenze culturali e temporali, con dinamiche non sempre lineari. Particolarmente rilevante è il cosiddetto *paradosso del consumo*, secondo cui in alcuni contesti una quota minoritaria di forti bevitori coesiste con un'ampia popolazione astinente.

Dal punto di vista metodologico, l'adozione di test non parametrici (Wilcoxon, Kruskal-Wallis, Friedman), motivati da test di normalità e QQ-plot, ha garantito la solidità e la robustezza statistica delle inferenze. Le visualizzazioni grafiche hanno facilitato l'interpretazione dei risultati e rafforzato le evidenze empiriche.

Sviluppi futuri Tra le possibili estensioni si segnalano:

- l'integrazione di nuove variabili, come orientamento religioso, istruzione o politiche sull'alcol;
- l'analisi dell'impatto di eventi esogeni (es. pandemia, conflitti);
- il collegamento con indicatori sanitari (mortalità, dipendenza);
- lo sviluppo di modelli predittivi e profili sintetici di rischio.

In conclusione, lo studio conferma l'utilità dei dati disaggregati WHO nell'individuare disuguaglianze strutturali e offre strumenti per supportare strategie di prevenzione e intervento basate su evidenze.

REFERENCES

- [1] World Health Organization, "Health Inequality Data Repository," 2024. [Online]. Available: <https://www.who.int/data/inequality-monitor/data>
- [2] World Health Organization, "Indicator Metadata: Alcohol Consumption Indicators," Health Inequality Data Repository, Oct. 2024. [Online]. Available: <https://www.who.int/data/inequality-monitor/data>
- [3] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965.
- [4] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945.
- [5] W. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *Journal of the American Statistical Association*, vol. 47, no. 260, pp. 583–621, 1952.
- [6] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *Journal of the American Statistical Association*, vol. 32, no. 200, pp. 675–701, 1937.
- [7] O. J. Dunn, "Multiple comparisons using rank sums," *Technometrics*, vol. 6, no. 3, pp. 241–252, 1964.
- [8] C. Bonferroni, "Teoria statistica delle classi e calcolo delle probabilità," *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, 1936.
- [9] C. Spearman, "The proof and measurement of association between two things," *The American Journal of Psychology*, vol. 15, no. 1, pp. 72–101, 1904.