



A.D. 1308
unipg

UNIVERSITÀ DEGLI STUDI
DI PERUGIA

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

Daniele Angeloni , Alessandra Ruggeri

Corso di Laurea Magistrale in Ingegneria Informatica
Curriculum Data Science and Data Engineering
Università degli Studi di Perugia

A.A. 2024/2025

Introduzione

Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica:
Verifica della normalità

Analisi statistica:
Test non parametrici

Conclusioni e sviluppi futuri



Table of Contents

Introduzione

Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica: Verifica della normalità

Analisi statistica: Test non parametrici

Conclusioni e sviluppi futuri

1 Introduzione

2 Dataset e metodologie

3 Exploratory Data Analysis (EDA)

4 Analisi statistica: Verifica della normalità

5 Analisi statistica: Test non parametrici

6 Conclusioni e sviluppi futuri



Disuguaglianze nel Consumo di Alcol

Introduzione

Dataset e
metodologie

Exploratory
Data Analysis
(EDA)

Analisi
statistica:
Verifica della
normalità

Analisi
statistica:
Test non
parametrici

Conclusioni e
sviluppi futuri

- Il consumo di alcol è un importante indicatore di disuguaglianza nella salute pubblica.
- Riflette differenze nei comportamenti sanitari tra gruppi socioeconomici, culturali e demografici.
- L'analisi si focalizza su:
 - Diffusione (% consumatori)
 - Consumo (litri medi tra i bevitori)
 - Astensione (% astinenti a vita)



Obiettivi del Progetto

Introduzione

Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica:
Verifica della normalità

Analisi statistica:
Test non parametrici

Conclusioni e sviluppi futuri

- Esplorare disuguaglianze nel consumo di alcol a livello globale.
- Analisi esplorativa dei dati disaggregati per:
 - Genere
 - Regione geografica (WHO)
 - Fascia di reddito (World Bank)
 - Anni (2000-2020)
- Verificare la significatività statistica delle differenze osservate.
- Supportare politiche sanitarie mirate e basate su evidenze.



Descrizione del Dataset

- **Fonte dati:** Health Inequality Data Repository – WHO Global Health Observatory (2024) [1] [2].
- **Copertura temporale:** 2000–2020 (per gli indicatori selezionati).
- **Unità di osservazione:** per paese, anno, genere, gruppo di reddito e regione geografica.
- **Indicatori analizzati:**
 - Alcohol, consumers past 12 months (%) — **Perc_Cons:** percentuale della popolazione (15+) che ha consumato alcol negli ultimi 12 mesi.
 - Alcohol, drinkers only per capita (15+) consumption (litres of pure alcohol) — **Litres_Cons:** litri medi di alcol puro consumati per bevitore (15+).
 - Alcohol, abstainers lifetime (%) — **Perc_Ast:** percentuale della popolazione (15+) che non ha mai consumato alcol nella vita.

Introduzione

Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica:
Verifica della normalità

Analisi statistica:
Test non parametrici

Conclusioni e sviluppi futuri



Metodologie Analitiche

Workflow in R:

- **Preprocessing:** Fase di pulizia, riorganizzazione e armonizzazione dei dati, con adeguamento dei nomi delle variabili e gestione dei valori mancanti.
- **Controllo qualità dati:** range validi, distribuzioni, outlier.
- **Analisi Esplorativa (EDA):** visualizzazioni aggregate, confronto tra gruppi, analisi temporali.
- **Test di normalità:** Shapiro-Wilk [3] e QQ-plot.
- **Test statistici:**
 - Wilcoxon rank-sum [4] (equivalente del Mann-Whitney) (confronto tra due gruppi indipendenti)
 - Kruskal-Wallis[5] (con tre o più gruppi indipendenti)
 - Friedman[6] (misure ripetute)
 - Test post-hoc di Dunn[7] (con correzione Bonferroni[8])

Introduzione

Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica: Verifica della normalità

Analisi statistica: Test non parametrici

Conclusioni e sviluppi futuri



Preprocessing e controllo qualità dei Dati

Prima dell'analisi esplorativa e statistica, è stata effettuata una fase di pulizia e validazione dei dati per garantirne l'affidabilità e la coerenza.

Preprocessing e controllo qualità dei dati:

- Rimozione delle osservazioni con valori NA (Not Available) significativi nella variabile `value` (numero esiguo).
- Selezione delle variabili rilevanti e ridenominazione dei nomi delle etichette per renderle più interpretabili (`country`, `year`, `gender`, `income_group`, `who_region`); e per gli indicatori: `Perc_Cons`, `Litres_Cons`, `Perc_Ast`.
- Riclassificazione dei gruppi di reddito mancanti con l'etichetta "Unknown", per preservare l'integrità dei dati.
- Verifica dei range e dei valori validi per ciascun indicatore/variabile e analisi preliminare delle distribuzioni e presenza di outlier.

Introduzione

Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica:
Verifica della normalità

Analisi statistica:
Test non parametrici

Conclusioni e sviluppi futuri



Exploratory Data Analysis

L'analisi esplorativa dei dati (EDA) ha l'obiettivo di:

- Investigare la struttura e la distribuzione dei tre indicatori selezionati:
 - Perc_Cons — % consumatori ultimi 12 mesi
 - Litres_Cons — litri di alcol puro per bevitore
 - Perc_Ast — % astinenti a vita
- Analizzare differenze per:
 - Genere
 - Regione WHO
 - Gruppo di reddito
 - Anno
- Individuare outlier, anomalie, trend e pattern regionali o demografici.

Le visualizzazioni seguenti offrono un primo sguardo descrittivo alla struttura del dataset WHO.

Introduzione

Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica: Verifica della normalità

Analisi statistica: Test non parametrici

Conclusioni e sviluppi futuri



A. Analisi generale

Distribuzioni dei valori dei tre indicatori principali (Litres_Cons, Perc_Cons, Perc_Ast) sull'intero campione, utili per confrontare scale e variabilità.

Introduzione

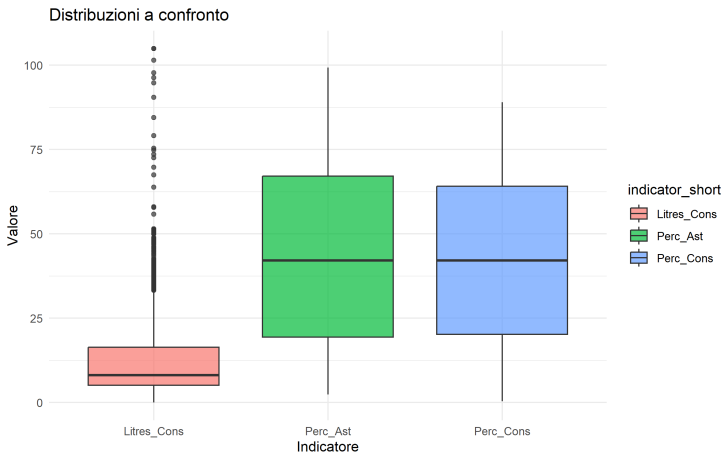
Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica:
Verifica della normalità

Analisi statistica:
Test non parametrici

Conclusioni e sviluppi futuri

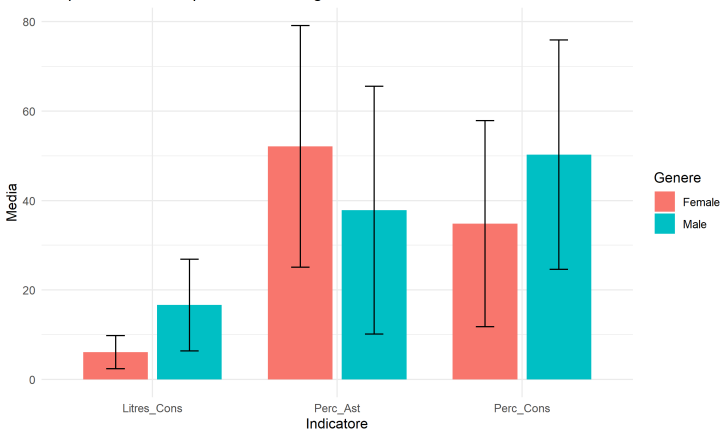




A. Analisi generale

Confronto tra media e deviazione standard dei tre indicatori suddivisi per genere. Evidenziate forti differenze tra uomini e donne.

Barplot media e SD per indicatore e genere



Introduzione

Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica: Verifica della normalità

Analisi statistica: Test non parametrici

Conclusioni e sviluppi futuri



Trend Globale per Indicatori (Male vs Female)

Andamento temporale aggregato a livello globale dei tre indicatori principali, disaggregati per genere. Mostrano una tendenza generale verso la diminuzione del consumo medio tra i bevitori e verso l'aumento della percentuale di consumatori.

Introduzione

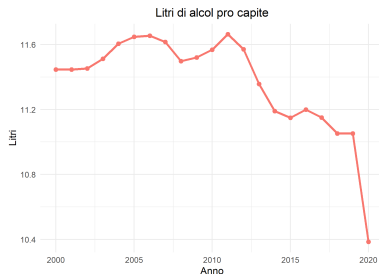
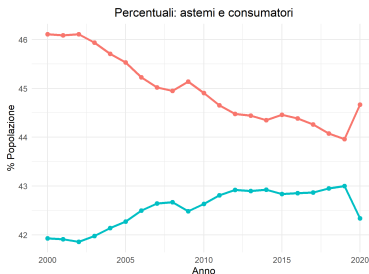
Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica:
Verifica della normalità

Analisi statistica:
Test non parametrici

Conclusioni e sviluppi futuri



Indicatore Perc_Ast Perc_Cons Litres_Cons



A. Analisi generale

Riepilogo generale dei dati: a sinistra la matrice di correlazione (Spearman [9]) tra i tre indicatori principali di consumo alcolico, evidenziando una forte correlazione negativa tra astinenza e consumo; a destra media e deviazione standard per genere e indicatore.

| Perc_Cons | Litres_Cons | Perc_Ast |
|-----------|-------------|----------|
| 1 | 0.34 | -0.98 |
| 0.34 | 1 | -0.33 |
| -0.98 | -0.33 | 1 |

| indicator_short | gender | media | sd |
|-----------------|--------|-------|-------|
| Litres_Cons | Female | 6.11 | 3.69 |
| Litres_Cons | Male | 16.63 | 10.26 |
| Perc_Ast | Female | 52.10 | 27.03 |
| Perc_Ast | Male | 37.85 | 27.71 |
| Perc_Cons | Female | 34.83 | 23.06 |
| Perc_Cons | Male | 50.26 | 25.66 |

Introduzione

Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica: Verifica della normalità

Analisi statistica: Test non parametrici

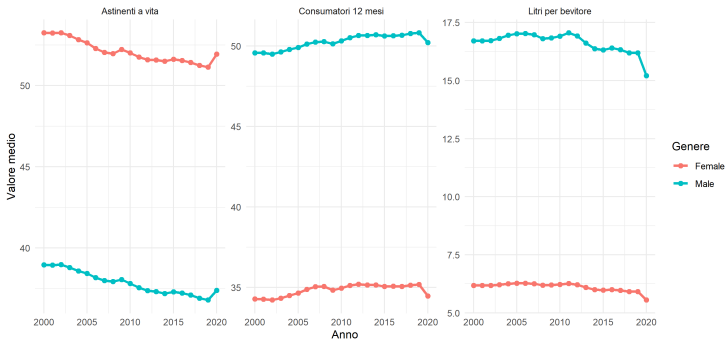
Conclusioni e sviluppi futuri



A. Analisi generale

Evoluzione nel tempo della media per ciascun indicatore, separata per genere. Permette di osservare andamenti distinti per uomini e donne.

Trend temporali per genere dei tre indicatori di consumo alcolico

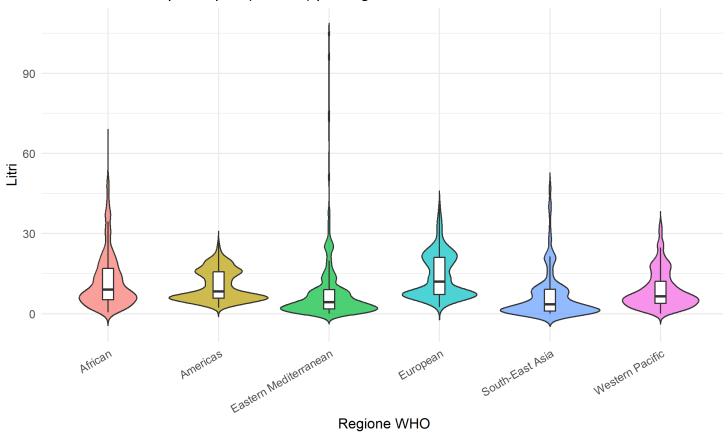




A. Analisi generale

Distribuzione dei litri medi di alcol puro per bevitore nelle diverse regioni WHO. Evidenti differenze culturali e geografiche.

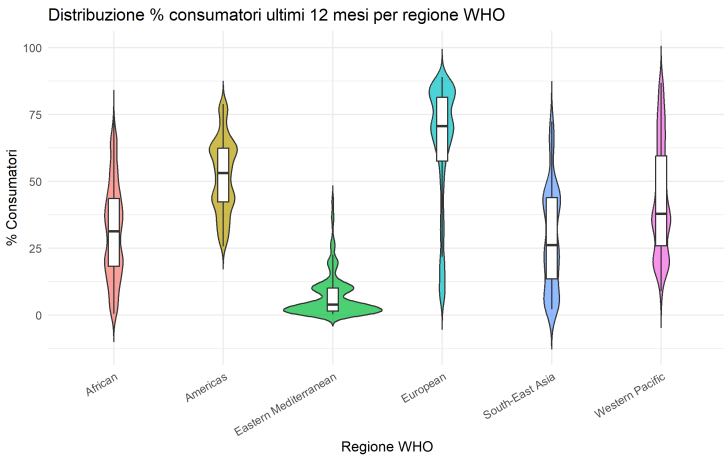
Distribuzione litri pro capite (bevitori) per regione WHO





A. Analisi generale

Distribuzione percentuale dei consumatori di alcol negli ultimi 12 mesi, suddivisa per regione WHO. Evidenziate aree a forte consumo.



Introduzione

Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica: Verifica della normalità

Analisi statistica: Test non parametrici

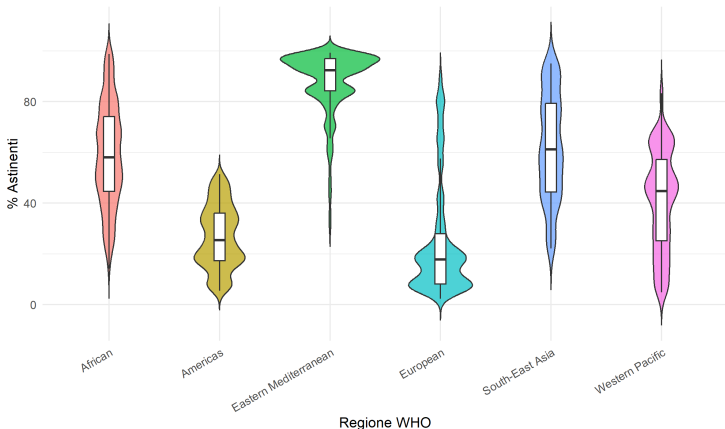
Conclusioni e sviluppi futuri



A. Analisi generale

Distribuzione della percentuale di popolazione che non ha mai consumato alcol, distinta per regione. Alcune aree mostrano valori estremi.

Distribuzione % astinenti a vita per regione WHO



Introduzione

Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica: Verifica della normalità

Analisi statistica: Test non parametrici

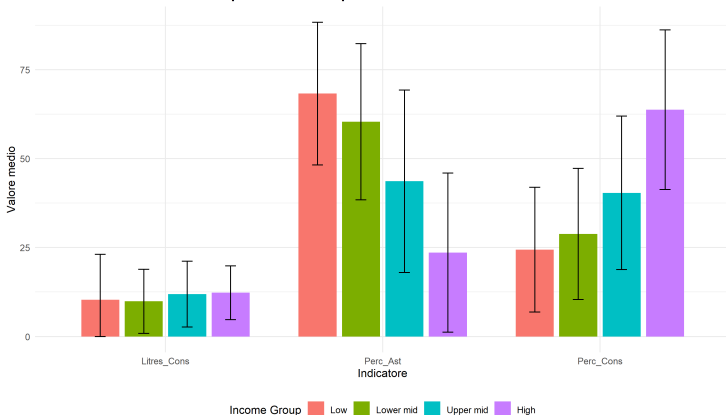
Conclusioni e sviluppi futuri



B. Analisi per fascia di reddito (Income Group)

Confronto tra media e deviazione standard dei tre indicatori per fasce di reddito. I paesi ad alto reddito mostrano maggiori consumi.

Barplot media e SD per indicatore e livello di reddito



Introduzione

Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica: Verifica della normalità

Analisi statistica: Test non parametrici

Conclusioni e sviluppi futuri



C. Variazioni annuali regionali

Variazioni annuali nella percentuale di **consumatori** e **astinenti a vita**, calcolate rispetto all'anno precedente. Le differenze evidenziano andamenti non omogenei tra le regioni WHO.

Introduzione

Dataset e metodologie

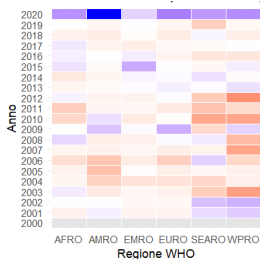
Exploratory Data Analysis (EDA)

Analisi statistica:
Verifica della normalità

Analisi statistica:
Test non parametrici

Conclusioni e sviluppi futuri

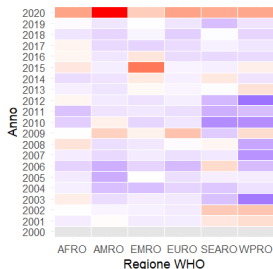
Consumatori



$\Delta\%$ rispetto all'anno precedente



Astinenti a vita



$\Delta\%$ astinenti rispetto anno prec.

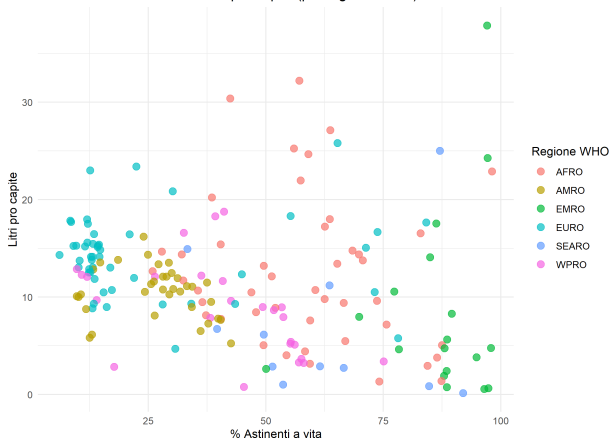




D. Anomalie e paradosso del consumo

Relazione tra percentuale di astinenti e consumo medio tra i bevitori.
Alcune regioni mostrano pattern estremi e paradossali.

Relazione tra % astinenti e litri pro capite (per regione WHO)

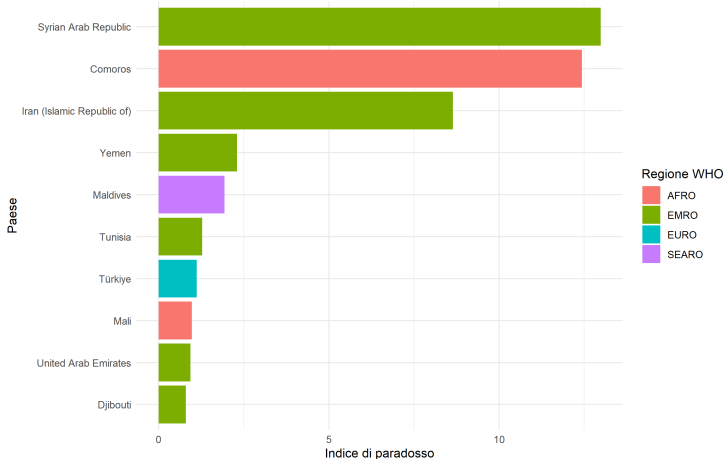




D. Anomalie e paradosso del consumo

I paesi con maggiore divario tra alta astinenza e alto consumo tra i bevitori. Spesso associati a contesti religiosi o culturali specifici.

Top 10 paesi con elevato indice paradosso: alto consumo + alta astinenza



Introduzione

Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica: Verifica della normalità

Analisi statistica: Test non parametrici

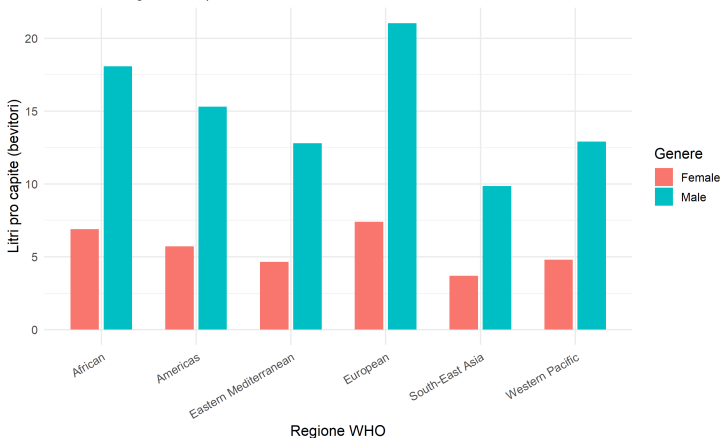
Conclusioni e sviluppi futuri



E. Differenze di Genere (Gender Gap)

Litri medi pro capite consumati da uomini e donne in ciascuna regione WHO. Gap confermato in ogni regione.

Litri medi pro capite per regione WHO e genere
Media su tutti gli anni disponibili



Introduzione

Dataset e
metodologie

Exploratory
Data Analysis
(EDA)

Analisi
statistica:
Verifica della
normalità

Analisi
statistica:
Test non
parametrici

Conclusioni e
sviluppi futuri

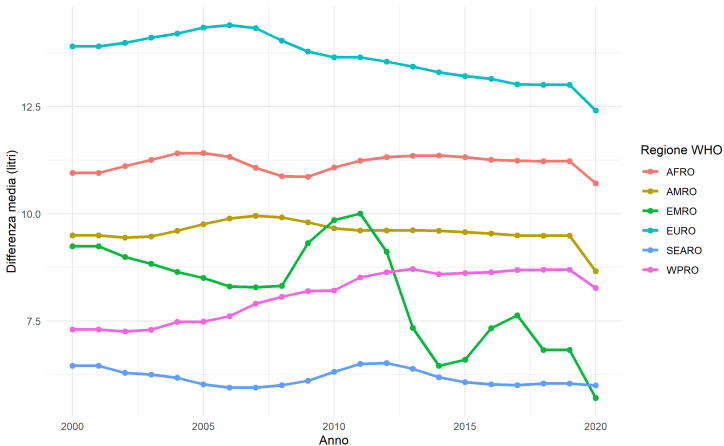


E. Differenze di Genere (Gender Gap)

Differenza media annua (uomini – donne) nel consumo tra bevitori.
Alcune regioni mostrano riduzioni, altre un divario stabile.

Gender gap (M – F) nei litri pro capite tra i bevitori – per regione WHO

Differenza media annuale: Male - Female





Verifica della Normalità

Introduzione

Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica:
Verifica della normalità

Analisi statistica:
Test non parametrici

Conclusioni e sviluppi futuri

- Prima di scegliere il test statistico appropriato, è necessario verificare l'assunzione di normalità della distribuzione dei dati.
- A tal fine, sono stati applicati:
 - **Shapiro-Wilk test**: verifica se un campione proviene da una distribuzione normale, fornendo evidenza statistica.
 - **QQ-plot (Quantile-Quantile plot)**: confronto grafico tra i quantili osservati e teorici di una distribuzione normale. Consente di visualizzare deviazioni dalla normalità.
- **Interpretazione:**
 - Se *p-value* minore di 0.05 \Rightarrow la normalità è respinta.
 - Se i punti nel QQ-plot si discostano dalla diagonale \Rightarrow possibile asimmetria o presenza di outlier.
- I risultati della verifica guideranno la scelta tra test parametrici e non parametrici nelle analisi successive.



Test di Normalità: Litres_Cons \sim gender

Introduzione

Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica:
Verifica della normalità

Analisi statistica:
Test non parametrici

Conclusioni e sviluppi futuri

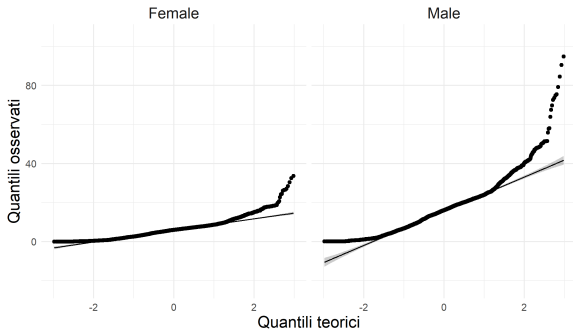


Table: Risultati del test Shapiro per Litres_Cons rispetto a gender

| gruppo | categoria | n | W | p-value |
|--------|-----------|------|-------|--------------|
| gender | Female | 3835 | 0.894 | 1.500000e-45 |
| gender | Male | 3835 | 0.897 | 4.170000e-45 |



Test di Normalità: Litres_Cons \sim income_group

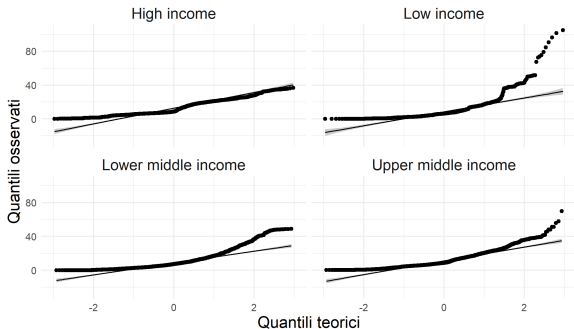


Table: Risultati del test Shapiro per Litres_Cons rispetto a income_group

| gruppo | categoria | n | W | p_value |
|--------------|---------------------|------|-------|--------------|
| income_group | Low income | 988 | 0.663 | 2.790000e-40 |
| income_group | Upper middle income | 2104 | 0.858 | 5.750000e-40 |
| income_group | High income | 2394 | 0.920 | 7.690000e-34 |
| income_group | Lower middle income | 2058 | 0.822 | 5.720000e-43 |



Test di Normalità: Perc_Ast \sim who_region

Introduzione

Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica:
Verifica della normalità

Analisi statistica:
Test non parametrici

Conclusioni e sviluppi futuri

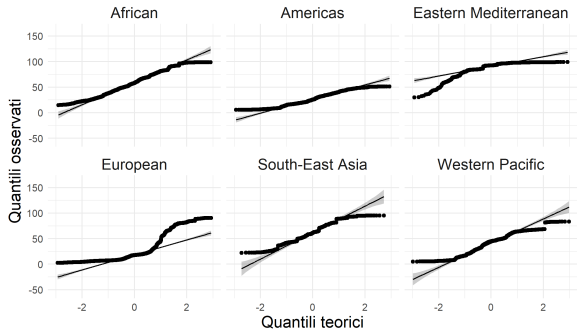


Table: Risultati del test Shapiro per Perc_Ast rispetto a who_region

| gruppo | categoria | n | W | p-value |
|------------|----------------|------|-------|----------|
| who_region | Eastern Medit. | 882 | 0.771 | 2.51e-33 |
| who_region | European | 2058 | 0.771 | 6.17e-47 |
| who_region | African | 1932 | 0.985 | 4.24e-13 |

| gruppo | categoria | n | W | p-value |
|------------|-----------------|------|-------|----------|
| who_region | Americas | 1470 | 0.967 | 8.96e-18 |
| who_region | Western Pacific | 1050 | 0.961 | 3.08e-16 |
| who_region | South-East Asia | 462 | 0.959 | 4.83e-10 |



Test di Normalità: Perc_Cons \sim income_group

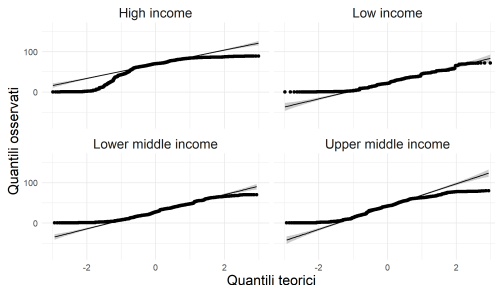


Table: Risultati del test Shapiro per Perc_Cons rispetto a income_group

| gruppo | categoria | n | W | p_value |
|--------------|---------------------|------|-------|----------|
| income_group | Low income | 1050 | 0.952 | 4.69e-18 |
| income_group | Upper middle income | 2100 | 0.960 | 1.52e-23 |
| income_group | High income | 2478 | 0.845 | 4.63e-44 |
| income_group | Lower middle income | 2100 | 0.962 | 5.52e-23 |



Test di Normalità: Perc_Cons \sim gender

Introduzione

Dataset e
metodologie

Exploratory
Data Analysis
(EDA)

Analisi
statistica:
Verifica della
normalità

Analisi
statistica:
Test non
parametrici

Conclusioni e
sviluppi futuri

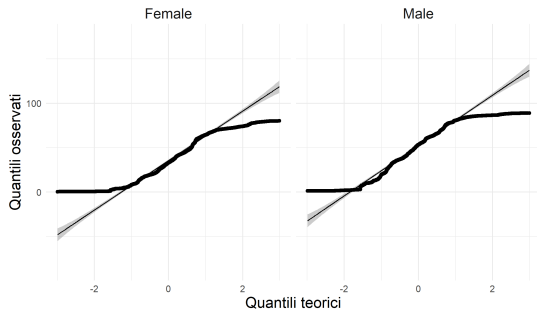


Table: Risultati del test Shapiro per Perc_Cons rispetto a gender

| gruppo | categoria | n | W | p_value |
|--------|-----------|------|-------|--------------|
| gender | Female | 3927 | 0.945 | 4.410000e-36 |
| gender | Male | 3927 | 0.948 | 2.570000e-35 |



Approccio ai Test Statistici non Parametrici

Introduzione

Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica:
Verifica della normalità

Analisi statistica:
Test non parametrici

Conclusioni e sviluppi futuri

Motivazione: I test di normalità (Shapiro–Wilk e QQ-plot) hanno evidenziato distribuzioni non normali. → Sono stati adottati test non parametrici, più robusti e appropriati in questi casi.

Test non parametrici applicati (basati su ranghi):

- **Wilcoxon rank-sum test (Mann-Whitney U test)** — confronto tra due gruppi indipendenti (es. gender)
- **Kruskal–Wallis** — confronto tra tre o più gruppi indipendenti (es. income group, who_region)
- **Friedman** — confronto tra misure ripetute su gruppi dipendenti (es. paesi nel tempo)

Post-hoc: Quando i test globali risultano significativi, si applica il **test di Dunn** con correzione Bonferroni (o equivalente per Friedman), per confronti a coppie.

Supporto grafico: Ogni test è accompagnato da un grafico (boxplot, line-plot ...) per facilitarne l'interpretazione visiva.



Wilcoxon rank-sum test: Litres_Cons \sim gender

Risultati del Wilcoxon rank-sum test:

Table: Risultati del Wilcoxon rank-sum test per
Wilcoxon_Litres_Cons_by_gender

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

Introduzione

Dataset e
metodologie

Exploratory
Data Analysis
(EDA)

Analisi
statistica:
Verifica della
normalità

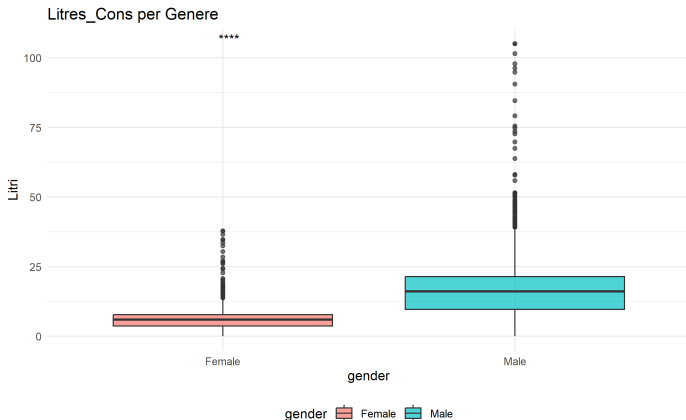
Analisi
statistica:
Test non
parametrici

Conclusioni e
sviluppi futuri



Confronto: Litres_Cons \sim gender

Boxplot comparativo per genere: confronto dei litri pro capite tra i bevitori.



Introduzione

Dataset e
metodologie

Exploratory
Data Analysis
(EDA)

Analisi
statistica:
Verifica della
normalità

Analisi
statistica:
Test non
parametrici

Conclusioni e
sviluppi futuri



Kruskal-Wallis: Litres_Cons \sim income_group

Risultati del test Kruskal per

Kruskal_Litres_

Cons_by_income_group

| Test | Statistica | df | p-value |
|----------------|------------|----|-----------|
| Kruskal-Wallis | 319.66 | 3 | < 2.2e-16 |

Post-hoc (Dunn + Bonferroni):

Table: Risultati del test Dunn per Dunn_Litres_Cons_by_income_group

| Comparison | Z | P_unadj | P_adj |
|---|------------|----------|----------|
| High income - Low income | 14.103927 | 3.59e-45 | 2.16e-44 |
| High income - Lower middle income | 14.249709 | 4.50e-46 | 2.70e-45 |
| Low income - Lower middle income | -2.712023 | 6.69e-03 | 4.01e-02 |
| High income - Upper middle income | 4.308836 | 1.64e-05 | 9.85e-05 |
| Low income - Upper middle income | -10.489657 | 9.64e-26 | 5.78e-25 |
| Lower middle income - Upper middle income | -9.663184 | 4.32e-22 | 2.59e-21 |

Introduzione

Dataset e
metodologie

Exploratory
Data Analysis
(EDA)

Analisi
statistica:
Verifica della
normalità

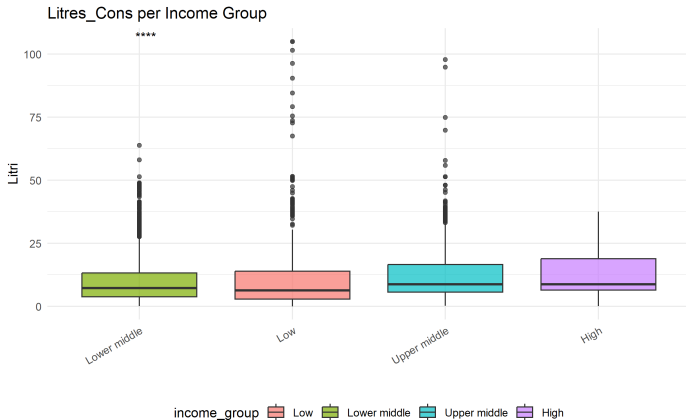
Analisi
statistica:
Test non
parametrici

Conclusioni e
sviluppi futuri



Confronto: Litres_Cons \sim income_group

Confronto dei consumi medi tra bevitori in base al livello di reddito.



Introduzione

Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica:
Verifica della normalità

Analisi statistica:
Test non parametrici

Conclusioni e sviluppi futuri



Kruskal-Wallis: Perc_Ast ~ who_region

Risultati del test Kruskal per

Kruskal_Perc_Ast

_by_who_region

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

Post-hoc (Dunn + Bonferroni):

Table: 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 |

| Comparison | Z | P_unadj | P_adj |
|--------------|------------|-----------|-----------|
| 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 |

Introduzione

Dataset e
metodologie

Exploratory
Data Analysis
(EDA)

Analisi
statistica:
Verifica della
normalità

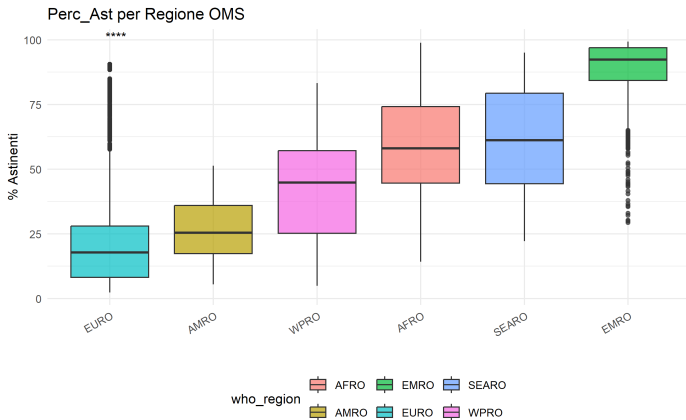
Analisi
statistica:
Test non
parametrici

Conclusioni e
sviluppi futuri



Confronto: Perc_Ast \sim who_region

Distribuzione delle percentuali di astinenti a vita nelle varie regioni.



Introduzione

Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica:
Verifica della normalità

Analisi statistica:
Test non parametrici

Conclusioni e sviluppi futuri



Kruskal-Wallis: Perc_Cons \sim income_group

Risultati test Kruskal-Wallis:

Table: Risultati del test Kruskal per Kruskal_Perc_Cons_by_income_group

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

Post-hoc (Dunn + Bonferroni):

Table: 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 |

Introduzione

Dataset e
metodologie

Exploratory
Data Analysis
(EDA)

Analisi
statistica:
Verifica della
normalità

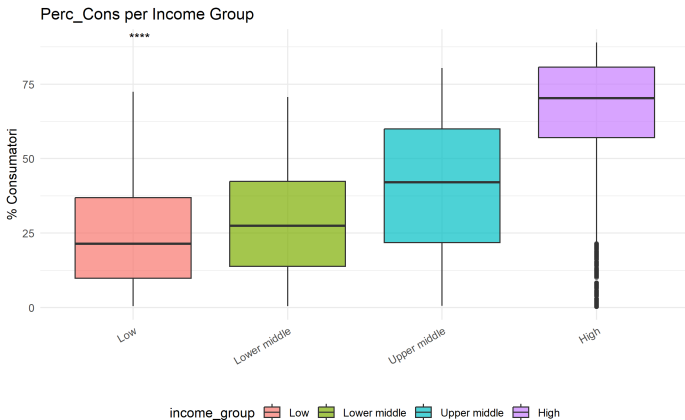
Analisi
statistica:
Test non
parametrici

Conclusioni e
sviluppi futuri



Confronto: Perc_Cons \sim income_group

Differenze significative nel tasso di consumo per livello di reddito.



Introduzione

Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica:
Verifica della normalità

Analisi statistica:
Test non parametrici

Conclusioni e sviluppi futuri



Wilcoxon rank-sum test: Perc_Cons \sim gender

Risultati del Wilcoxon rank-sum test:

Table: Risultati del Wilcoxon rank-sum test per
Wilcoxon_Perc_Cons_by_gender

| Test | Statistica | df | p-value |
|----------|------------|----|-------------|
| Wilcoxon | 5066090 | - | $< 2.2e-16$ |

Introduzione

Dataset e
metodologie

Exploratory
Data Analysis
(EDA)

Analisi
statistica:
Verifica della
normalità

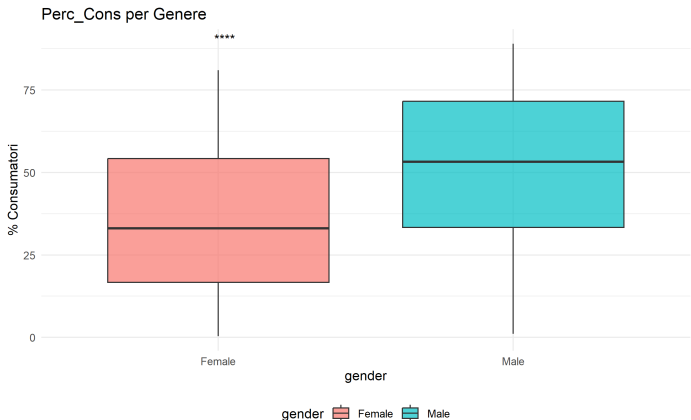
Analisi
statistica:
Test non
parametrici

Conclusioni e
sviluppi futuri



Confronto: Perc_Cons \sim gender

Boxplot percentuale consumatori per genere: confronto statistico e visivo.



Introduzione

Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica: Verifica della normalità

Analisi statistica: Test non parametrici

Conclusioni e sviluppi futuri



Friedman Test: Litres_Cons \sim year (Italy, France, Germany)

Introduzione

Dataset e
metodologie

Exploratory
Data Analysis
(EDA)

Analisi
statistica:
Verifica della
normalità

Analisi
statistica:
Test non
parametrici

Conclusioni e
sviluppi futuri

Risultati del test Friedman (2000–2020) per

`friedman_litres_italy`
`_france_germany`

| Test | Statistica | df | p-value |
|----------|------------|----|-----------|
| Friedman | 47.031 | 20 | 0.0005809 |

Post-hoc (Friedman + Bonferroni):

Table: 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 |

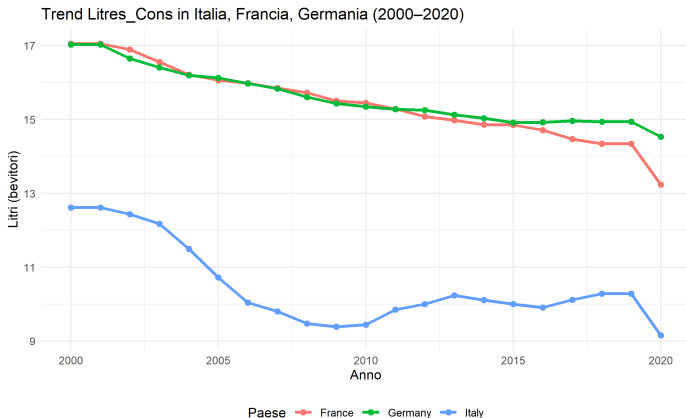
| Group1 | Group2 | p-value |
|--------|--------|----------|
| 2020 | 2002 | 0.005238 |
| 2020 | 2003 | 0.028373 |
| 2020 | 2004 | 0.105238 |
| 2020 | 2005 | 0.307619 |

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



Trend Litres_Cons – Italy, France, Germany (2000–2020)

Andamento comparativo nel tempo del consumo tra i bevitori in Italia, Francia e Germania.



Introduzione

Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica:
Verifica della normalità

Analisi statistica:
Test non parametrici

Conclusioni e sviluppi futuri



Conclusioni – Disuguaglianze strutturali

L'analisi ha evidenziato disuguaglianze marcate nel consumo di alcol a livello globale:

- Gli uomini risultano consumatori più assidui e intensi rispetto alle donne in tutte le regioni WHO.
- Nei paesi ad alto reddito prevale il consumo; nei paesi a basso reddito è più diffusa l'astinenza a vita.
- I livelli di consumo e di consumatori di alcol differiscono in modo sostanziale tra le regioni WHO, riflettendo fattori culturali, normativi e socio-economici.
- Il 2020 segna una discontinuità nei trend dei tre indicatori, con variazioni significative in molte regioni — verosimilmente associate agli effetti della pandemia COVID-19.

Un caso emblematico è il **paradosso del consumo**: in alcune regioni una minoranza di forti bevitori coesiste con un'elevata quota di astinenti, evidenziando polarizzazioni comportamentali.

Introduzione

Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica:
Verifica della normalità

Analisi statistica:
Test non parametrici

Conclusioni e sviluppi futuri



Conclusioni – Robustezza metodologica

Dal punto di vista statistico:

- I test di normalità (Shapiro–Wilk) hanno escluso la distribuzione normale nella maggior parte dei gruppi.
- I QQ-plot hanno supportato visivamente le deviazioni osservate, evidenziando asimmetrie e outlier.
- I test non parametrici (Wilcoxon rank-sum / Mann Whitney, Kruskal-Wallis e Friedman) hanno garantito inferenze robuste anche in presenza di distribuzioni eterogenee.
- L'integrazione con visualizzazioni grafiche ha rafforzato l'interpretabilità dei risultati.

Tutto il processo è stato realizzato in linguaggio **R**, sfruttando la flessibilità della pipeline di analisi (preprocessing, EDA, test inferenziali).

Introduzione

Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica: Verifica della normalità

Analisi statistica: Test non parametrici

Conclusioni e sviluppi futuri



Sviluppi futuri

L'analisi può essere ampliata in diverse direzioni:

- Integrare nuove variabili: *istruzione, religione, politiche sull'alcol.*
- Analizzare l'impatto di eventi esogeni: *pandemie, conflitti, cambi normativi.*
- Collegare il consumo a indicatori di salute pubblica: *mortalità, dipendenza, ospedalizzazioni.*
- Sviluppare modelli predittivi o profili sintetici di rischio per target specifici.

Queste estensioni potranno supportare strategie di prevenzione basate su evidenze, utilizzando al meglio i dati disaggregati WHO.

Introduzione

Dataset e metodologie

Exploratory Data Analysis (EDA)

Analisi statistica: Verifica della normalità

Analisi statistica: Test non parametrici

Conclusioni e sviluppi futuri



Riferimenti I

Riferimenti
bibliografici

- [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.