# ANALISI ESPLORATIVA DEL CONSUMO DI PRODOTTI A BASE DI TABACCO NEL MONDO DAL 2010 AL 2020

Data Science for Health Systems

Marco Venturi
Dipartimento di Ingegneria
Università degli studi di Perugia

# INTRODUZIONE

- L'epidemia del tabacco rappresenta una delle più grandi minacce per la salute pubblica che il mondo abbia mai affrontato, causando la morte di oltre 8 milioni di persone all'anno in tutto il mondo.

- Più di 7 milioni di queste morti sono il risultato dell'uso diretto del tabacco, mentre circa 1,3 milioni sono il risultato dell'esposizione al fumo passivo da parte dei non fumatori.

- Lo studio propone di analizzare il consumo di tabacco e prodotti a base di tabacco nel mondo nel periodo compreso tra il 2000 e il 2020, esaminando le differenze nell'incidenza tra diverse aree geografiche e tra i sessi.

## OBIETTIVO DELL'ANALISI

Questa analisi mira a comprendere meglio l'entità del problema del tabacco e a identificare le disparità geografiche e di genere nell'uso del tabacco durante il periodo 2010-2020 .

# IL DATASET

## DESCRIZIONE

- Il dataset analizzato è chiamato "Non-age-standardized estimates of current tobacco use, tobacco smoking, and cigarette smoking (Tobacco Control: Monitor)." Questi dati sono forniti dall'Organizzazione Mondiale della Sanità.
- Questo dataset contiene informazioni sul consumo di prodotti a base di tabacco.
- Il dataset fornisce informazioni sulla percentuale di consumatori nella popolazione in ciascun paese.
- Questi valori derivano dalla popolazione di età pari o superiore a 15 anni che attualmente utilizza qualsiasi prodotto a base di tabacco.
- Il dataset originale è composto da 13.284 campioni, ciascuno con 34 variabili.

# IL DATASET

## MODELLAZIONE

1. Rimozione delle variabili NaN e delle colonne non informative
2. Cambio nome delle variabili selezionate
3. Rimozione dell'intervallo di confidenza dalla variabile 'value'
4. Rimozione dei campioni raccolti nei anni 2023 e 2025
5. Riduzione dell'informazione ridondante

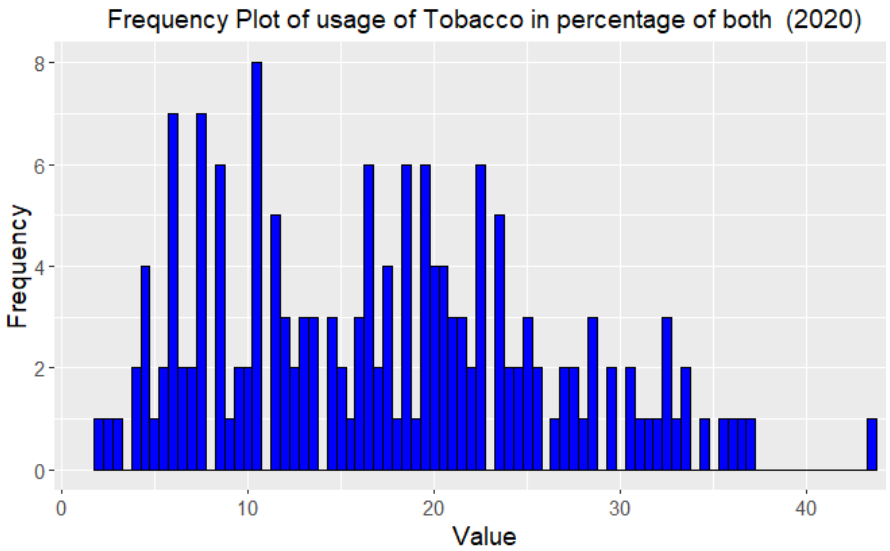
# IL DATASET

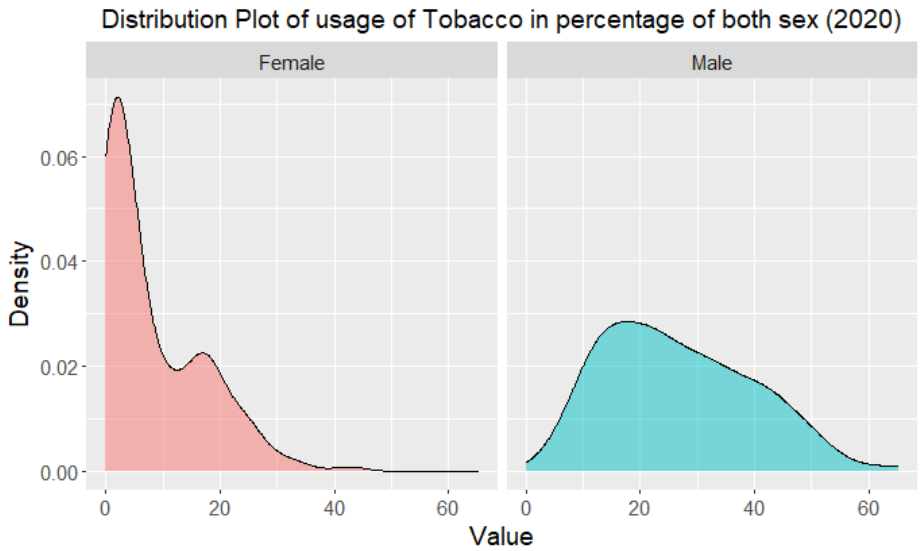
## DATASET FINALE

| | geo_region | state | year | sex | value |
|---|---|---|---|---|---|
| 3050 | Americas | Canada | 2020 | Male | 14 |
| 3051 | Africa | Mali | 2020 | Male | 15 |
| 3052 | Europe | Norway | 2020 | Female | 15 |
| 3053 | Africa | Eritrea | 2020 | Male | 15 |
| 3054 | Western Pacific | Australia | 2020 | Male | 15 |
| 3055 | Americas | Belize | 2020 | Male | 15 |
| 3056 | Africa | Eswatini | 2020 | Male | 15 |
| 3057 | Western Pacific | Tonga | 2020 | Female | 15 |
| 3058 | Europe | Portugal | 2020 | Female | 15 |
| 3059 | South-East Asia | Bangladesh | 2020 | Female | 15 |
| 3060 | Americas | Jamaica | 2020 | Male | 15 |
| 3061 | Africa | Burundi | 2020 | Male | 15 |
| 3062 | Europe | Norway | 2020 | Both sexes | 16 |

# ANALISI ESPLORATIVA

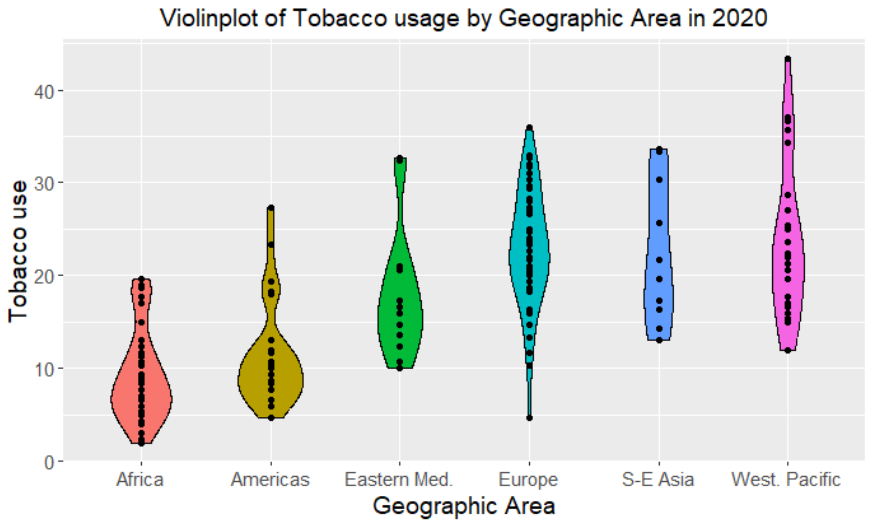
## 1. Distribuzione dei valori

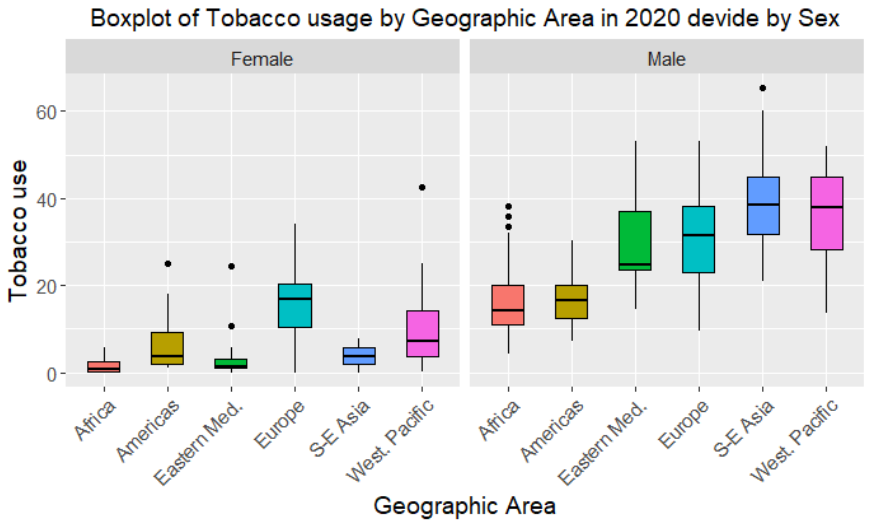


# ANALISI ESPLORATIVA

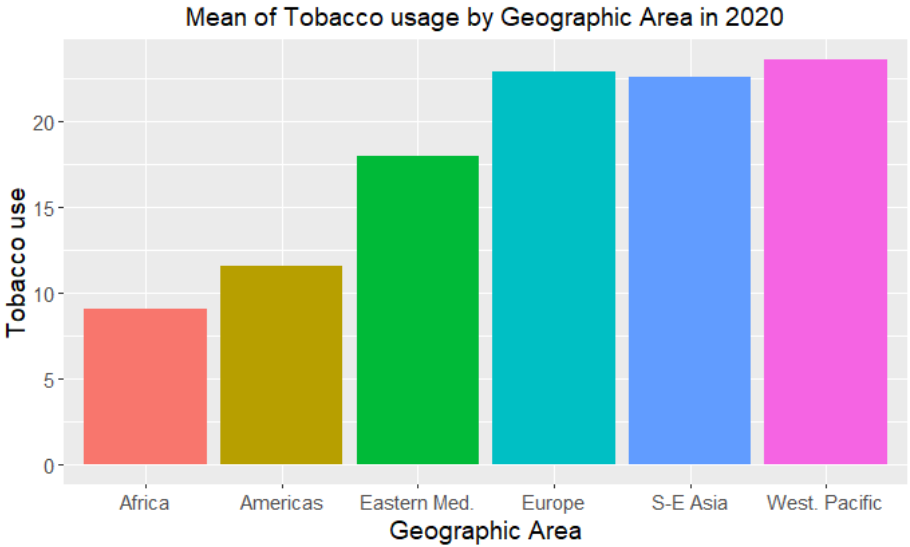
## 2. Distribuzione dei valori su area geografica

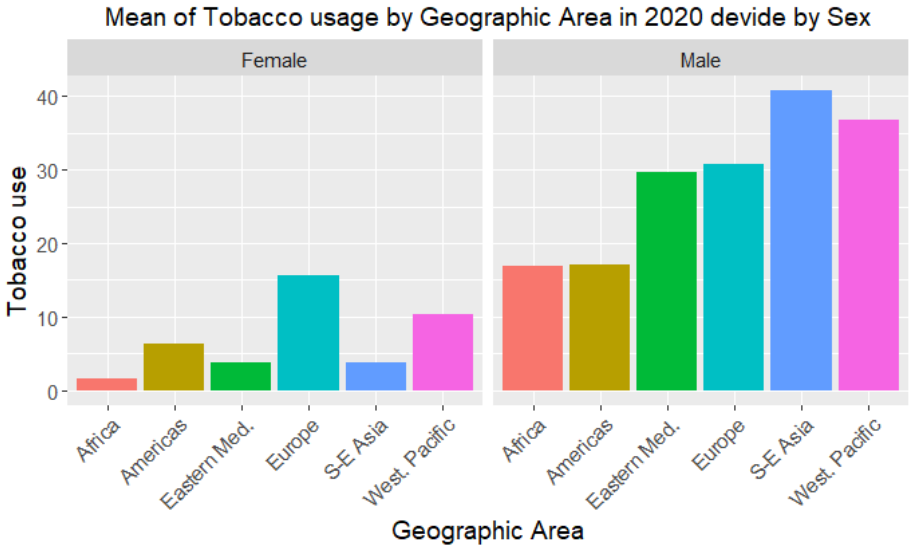


# ANALISI ESPLORATIVA

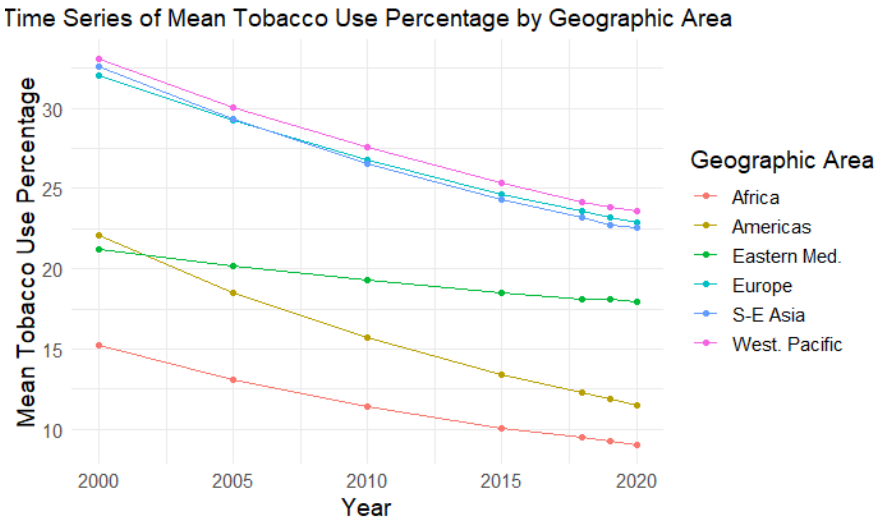
## 3. Valori delle medie per area geografica

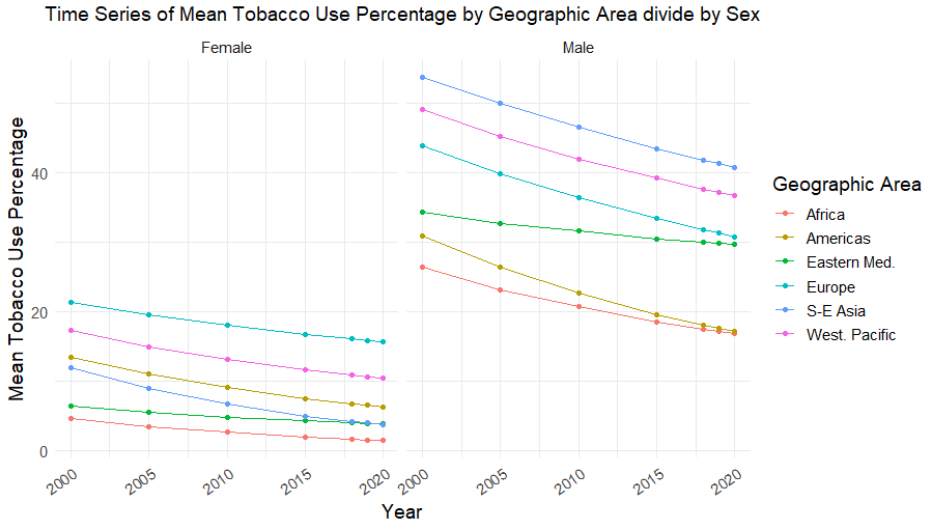


# ANALISI ESPLORATIVA

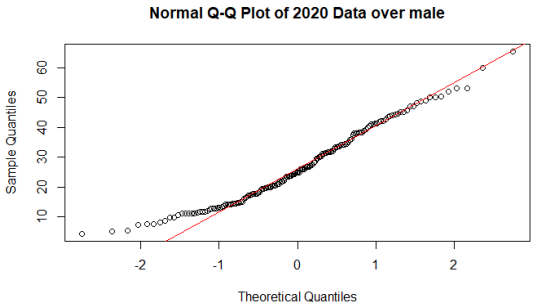
## 4. Valori delle medie delle aree geografiche nel tempo

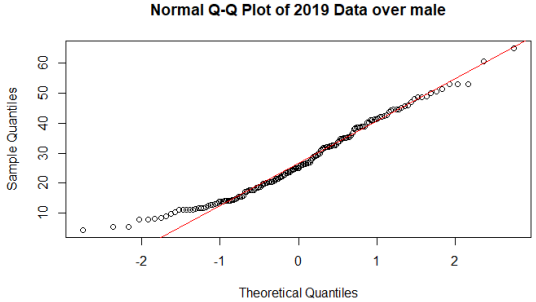


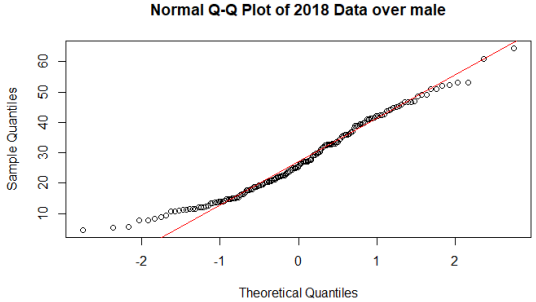



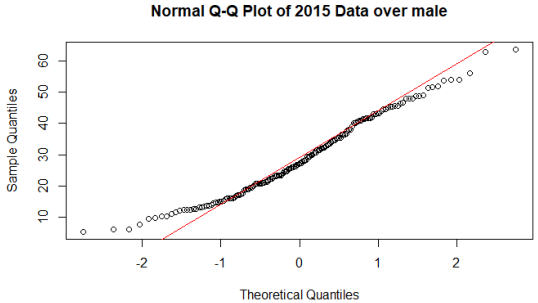
# TEST STATISTICI (Normalità)

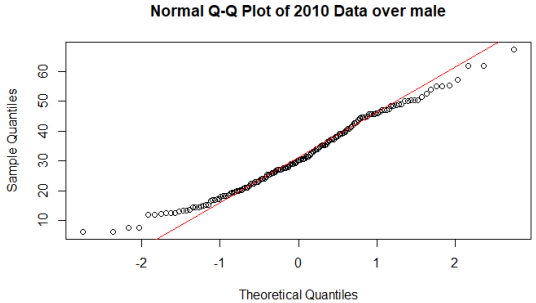
## 1. Plot del quantile teorico dei maschi



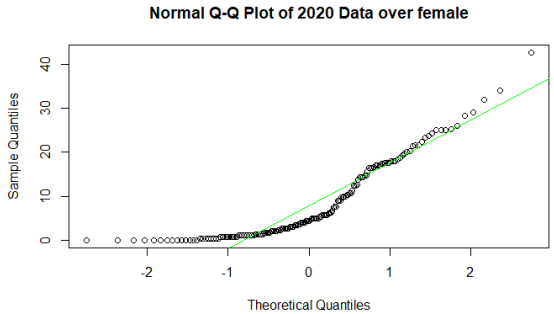



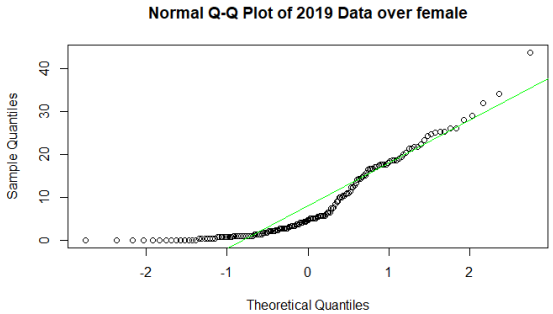



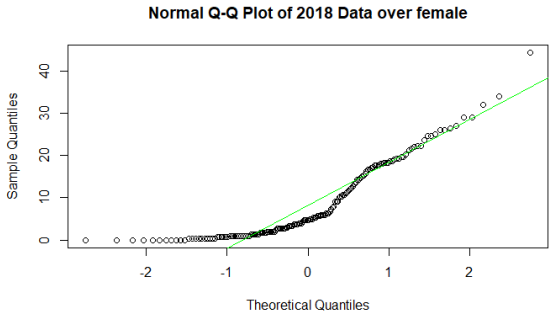



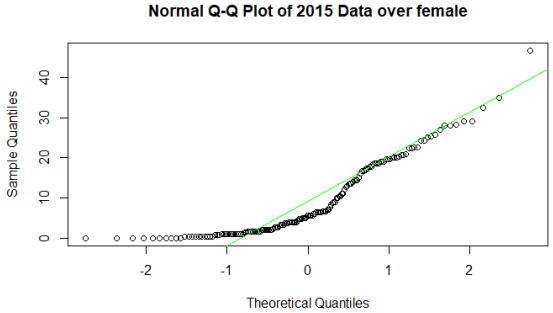
# TEST STATISTICI (Normalità)

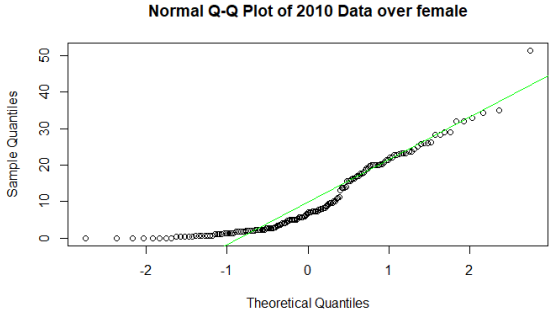
## 2. Plot del quantile teorico delle femmine



# TEST STATISTICI (Normalità)

## 3. Verifica della Normalità maschi

| Anni | Statistica W | P-value |
|---|---|---|
| 2010 | 0.98049 | 0.02058 |
| 2015 | 0.97356 | 0.003125 |
| 2018 | 0.96975 | 0.00118 |
| 2019 | 0.96925 | 0.001042 |
| 2020 | 0.96742 | 0.0006639 |

La normalità è rifiutata in tutti i gruppi di sesso maschile.

# TEST STATISTICI (Normalità)

## 4. Verifica della Normalità femmine

| Anni | Statistica W | P-value |
|---|---|---|
| 2010 | 0.86804 | 7.991x10^-11 |
| 2015 | 0.85454 | 1.827x10^-11 |
| 2018 | 0.84469 | 6.587x10^-12 |
| 2019 | 0.84195 | 4.998x10^-12 |
| 2020 | 0.83917 | 3.788x10^-12 |

La normalità è rifiutata in tutti i gruppi di sesso femminile.

# TEST STATISTICI (prerequisiti Anova)

## 1. Verifica dell'omoschedasticità (Test di Bartlett)

| Sesso | Statistica K-square | P-value |
|---|---|---|
| Maschio | 0.055455 | 0.9996 |
| Femmina | 3.431 | 0.4885 |

In entrambi i sessi è presente l'omoschedasticità.

# TEST STATISTICI (prerequisiti Anova)

## 2. Verifica della sfericità

| Sesso | Epsilon |
|---|---|
| Maschio | 0.9860699 |
| Femmina | 0.9866401 |

In entrambi i sessi il valore di epsilon non si discosta di troppo da 1 quindi la sfericità non influenza il risultato.

# TEST STATISTICI(ipotesi nulla)

## 1. Test dell'ipotesi su campioni maschi

- Si è deciso di applicare sia il test non parametrico di Friedman che il test di Anova.
- Questa decisione è dovuta al fatto che nonostante il test di shapiro-wilk rifiuti la normalità, la forma della distribuzione dei dati in questione è vicina alla forma della distribuzione normale, e i dati sembrano rispettare anche il pattern del quantile teorico.
- Anova e Friedman convergono a un risultato vicino.
- Il test F subjects invece risulta avere un valore nullo.

| Test | p-value |
|---|---|
| Anova | < 2 × 10^−16 |
| F subjects | 0 |
| Friedman | < 2.2 ×10−16 |

# TEST STATISTICI(ipotesi nulla)

## 2. Test dell’ipotesi su campioni Femminili

- Si è applicato il test non parametrico di Friedman sui campioni Femminili, poiché la normalità è rifiutata dal test di shapiro-wilk.
- Poiché entrambi i campioni di sesso maschile e femminile hanno confermato le ipotesi nulle, si prosegue con l’analisi post-hoc per osservare se la conferma dell’ipotesi nulla avviene anche tra i gruppi.

| Test | p-value |
|---|---|
| Friedman | < 2.2 ×10^−16 |

# TEST STATISTICI(analisi post-hoc)

## 1. Analisi post-hoc maschile

Correzione Bonferroni

| Anno 1 | Anno 2 | p-value |
|---|---|---|
| 2015 | 2010 | <2×10^−16 |
| 2018 | 2010 | <2×10^−16 |
| 2018 | 2015 | <2×10^−16 |
| 2019 | 2010 | <2×10^−16 |
| 2019 | 2015 | <2×10^−16 |
| 2019 | 2018 | <2.8×10^−15 |
| 2020 | 2010 | <2×10^−16 |
| 2020 | 2015 | <2×10^−16 |
| 2020 | 2018 | <2×10^−16 |
| 2020 | 2019 | <2×10^−16 |

Procedura Benjamini-Hochber

| Anno 1 | Anno 2 | p-value |
|---|---|---|
| 2015 | 2010 | <2×10^−16 |
| 2018 | 2010 | <2×10^−16 |
| 2018 | 2015 | <2×10^−16 |
| 2019 | 2010 | <2×10^−16 |
| 2019 | 2015 | <2×10^−16 |
| 2019 | 2018 | <2.8×10^−15 |
| 2020 | 2010 | <2×10^−16 |
| 2020 | 2015 | <2×10^−16 |
| 2020 | 2018 | <2×10^−16 |
| 2020 | 2019 | <2×10^−16 |

# TEST STATISTICI(analisi post-hoc)

## 1. Analisi post-hoc femmine

Correzione Bonferroni

| Anno 1 | Anno 2 | p-value |
|---|---|---|
| 2015 | 2010 | <2×10^−16 |
| 2018 | 2010 | <2×10^−16 |
| 2018 | 2015 | <1.1×10^−15 |
| 2019 | 2010 | <2×10^−16 |
| 2019 | 2015 | <2×10^−16 |
| 2019 | 2018 | <1.9×10^−10 |
| 2020 | 2010 | <2×10^−16 |
| 2020 | 2015 | <2×10^−16 |
| 2020 | 2018 | <6.1×10^−15 |
| 2020 | 2019 | <1.2×10^−8 |

Procedura Benjamini-Hochber

| Anno 1 | Anno 2 | p-value |
|---|---|---|
| 2015 | 2010 | <2×10^−16 |
| 2018 | 2010 | <2×10^−16 |
| 2018 | 2015 | <2×10^−16 |
| 2019 | 2010 | <2×10^−16 |
| 2019 | 2015 | <2×10^−16 |
| 2019 | 2018 | <2.2×10^−11 |
| 2020 | 2010 | <2×10^−16 |
| 2020 | 2015 | <2×10^−16 |
| 2020 | 2018 | <7.6×10^−16 |
| 2020 | 2019 | <1.2×10^−9 |

# CONCLUSIONI

Le analisi effettuate mostrano come l'andamento negli ultimi dieci anni sia in gruppi maschili che femminili abbia un trend decrescente.

Il Test post-hoc conferma come questo trend sia confermato anche in singoli anni e non solo nell'insieme.

# REFERENZE

- Il codice sorgente e la documentazione sono disponibili al seguente link:
  https://github.com/Arcaici/TobaccoControlMonitor_WHO

- Fonte dati → *Tobacco control: Monitor, World Health Organization*:
  https://www.who.int/data/gho/data/indicators/indicator-details/GHO/gho-tobacco-control-monitor-current-tobaccouse-tobaccosmoking-cigarrettesmoking-nonagestd-tobnonagestdcurr