

# ANALISI ESPLORATIVA DEL TASSO DI SUICIDIO ANNUALE IN DIVERSI PAESI SU DATASET WHO

TESINA DI DATA SCIENCE FOR HEALTH SYSTEMS

PAOLO SPEZIALI

INGEGNERIA INFORMATICA E ROBOTICA  
CURRICULUM DATA SCIENCE

A.A. 2022/2023



A.D. 1308  
**unipg**  
DIPARTIMENTO  
DI INGEGNERIA

# **INTRODUZIONE**

- **Suicidio:** atto intenzionale di terminare la propria vita.
- Il suicidio è una delle principali cause di morte nel mondo occidentale.
- Dati stabili a livello nazionale nel breve termine, predizione basata sui tassi passati.
- Studio sui cambiamenti globali nel tasso di suicidio (2000-2019) e confronto tra regioni del mondo e gruppi di paesi.

**DATASET**

## DESCRIZIONE DEL DATASET

- Origine del dataset: World Health Organization (WHO).
- Titolo del dataset: "Age-standardized suicide rates (per 100,000 population)".
- Copre il periodo dal 2000 al 2019 e include dati per soggetti maschili, femminili e di entrambi i sessi.
- Standardizzazione per la distribuzione dell'età della popolazione del paese.
- Acquisizione dei dati da censimenti, registrazioni anagrafiche e analisi di certificati medici.
- Struttura originale: 10'980 campioni, 34 feature, molte con valori nulli o non rilevanti per l'analisi.

# MODELLAZIONE DEL DATASET

Nome Originale	Nuovo nome
ParentLocation	WorldRegion
SpatialDimValueCode	CountryCode
Location	Country
Period	Year
Dim1	Sex
FactValueNumeric	Value

- Successivamente, il dataset è stato ulteriormente modificato per avere un'unica entry per ogni anno relativo a un paese.
- La feature "Sex" è stata suddivisa in tre diverse feature (Male, Female, Both), ognuna contenente il valore associato nel campo "Value."

# STRUTTURA FINALE DEL DATASET

Il nostro dataset è ora composto da 3660 campioni e 7 feature.

- WorldRegion: area geografica secondo la WHO.
- CountryCode: codice a tre lettere maiuscole del paese.
- Country: nome del paese.
- Year: anno del dato.
- Male: tasso di suicidi maschile.
- Female: tasso di suicidi femminile.
- Both: tasso di suicidi di entrambi i sessi.

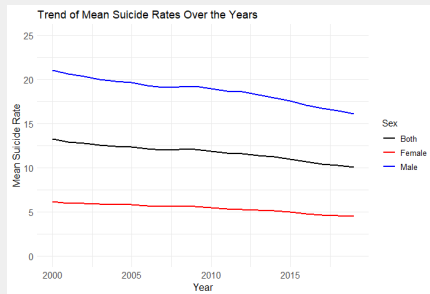
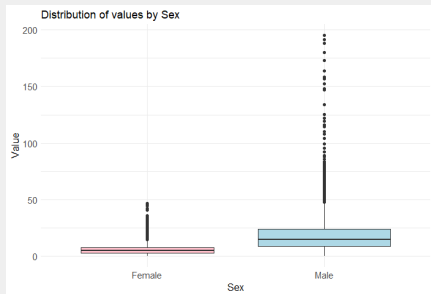
# **ANALISI ESPLORATIVA DEI DATI (EDA)**



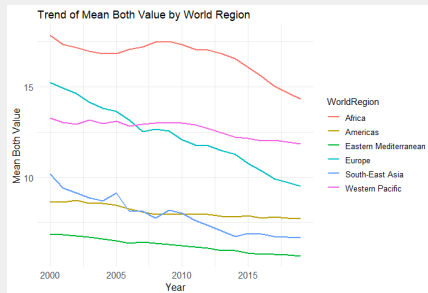
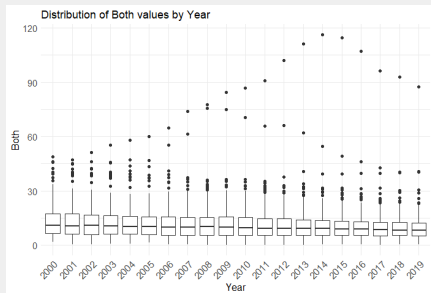
Faremo l'analisi dell'intero dataset e poi di quattro gruppi di soggetti a seconda dei loro valori Both per ogni anno (2014-2019):

1. Regioni globali (WR)
2. Paese con PIL più alto per WR
3. Paesi con tasso di suicidio più alto in assoluto
4. Paesi con tasso di suicidio più alto per WR

# IL QUADRO GENERALE - DISTRIBUZIONE PER SESSO

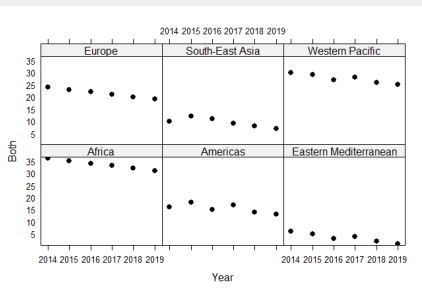


# IL QUADRO GENERALE - DISTRIBUZIONE PER ANNO

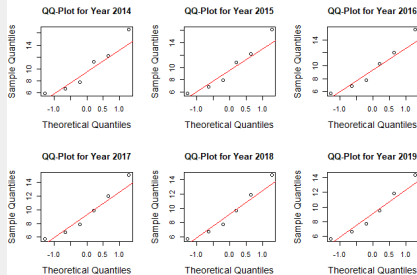


# PRIMO GRUPPO - REGIONI GLOBALI

- Africa
- Eastern Mediterranean
- Americas



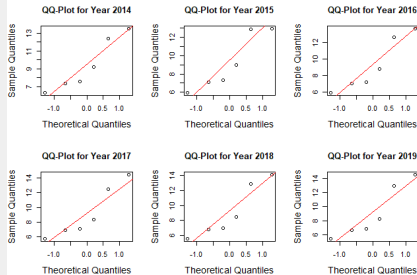
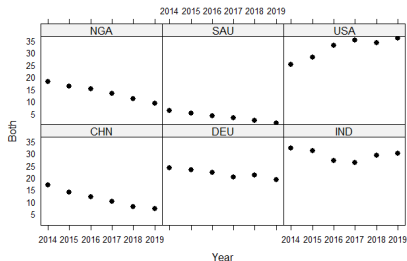
- Europe
- Western Pacific
- South-East Asia



# SECONDO GRUPPO - PAESE CON PIL PIÙ ALTO PER WR

- **Nigeria** (NGA) per Africa
- **Saudi Arabia** (SAU) per Eastern Mediterranean
- **United States of America** (USA) per Americas

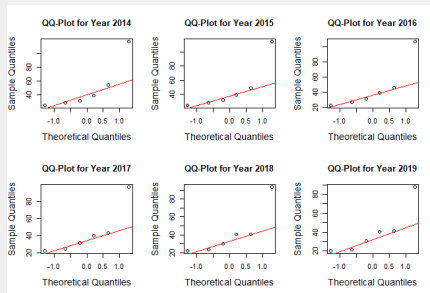
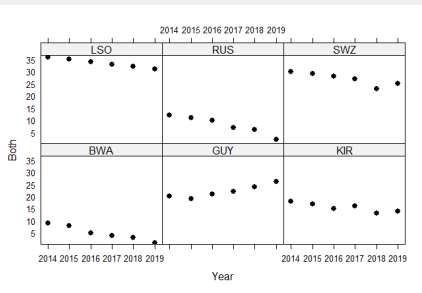
- **Germany** (DEU) per Europe
- **China** (CHN) per Western Pacific
- **India** (IND) per South-East Asia



# TERZO GRUPPO - PAESI CON TASSO DI SUICIDIO PIÙ ALTO IN ASSOLUTO

- Lesotho (LSO)
- Russian Federation (RUS)
- Eswatini (SWZ)

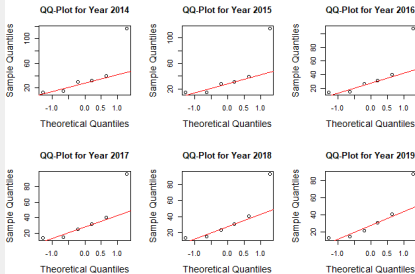
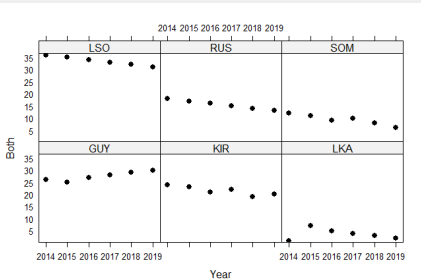
- Botswana (BWA)
- Guyana (GUY)
- Kiribati (KIR)



# QUARTO GRUPPO - PAESI CON TASSO DI SUICIDIO PIÙ ALTO PER WR

- **Lesotho (LSO)** per Africa
- **Russian Federation (RUS)** per Europe
- **Somalia (SOM)** per Eastern Mediterranean

- **Guyana (GUY)** per Americas
- **Kiribati (KIR)** per Western Pacific
- **Sri Lanka (LKA)** per South-East Asia



# TEST STATISTICI



# OBIETTIVO DEI TEST STATISTICI

- **Obiettivo:** Verificare che i dati, all'interno dei vari gruppi, provengano da distribuzioni che non differiscono in maniera significativa tra di loro.
- **Strumento:** ANOVA a misure ripetute.
- **Presupposti:**
  - ▶ Verifica della normalità
  - ▶ Verifica dell'omocedasticità
  - ▶ Verifica della sfericità
- Se una di queste assunzioni non è rispettata, si usa il test non parametrico di Friedman.

# VERIFICA DELLA NORMALITÀ

Test di Shapiro-Wilk

Gruppo	Anni					
	2014	2015	2016	2017	2018	2019
<b>1</b>	0.531	0.672	0.690	0.705	0.708	0.719
<b>2</b>	0.913	0.824	0.804	0.832	0.746	0.569
<b>3</b>	0.018	0.011	0.017	0.026	0.021	0.056
<b>4</b>	0.015	0.011	0.018	0.037	0.046	0.071

Solo i primi due gruppi rispettano il vincolo della normalità.

# VERIFICA DELL'OMOSCHEDASTICITÀ

Test di Bartlett

Gruppo	K-squared	p-value
<b>1</b>	0.27014	0.9982
<b>2</b>	0.39744	0.9954

Test di Levene

Gruppo	F-value	p-value
<b>3</b>	0.037	0.9992
<b>4</b>	0.016	0.9999

Tutti i gruppi rispettano il vincolo dell'omoschedasticità.

# VERIFICA DELLA SFERICITÀ

## Statistica di Greenhouse-Geisser

<b>Gruppo</b>	$\epsilon$
<b>1</b>	0.2067
<b>2</b>	0.2280
<b>3</b>	0.2122
<b>4</b>	0.2028

Nessuno dei quattro gruppi soddisfa la sfericità, i loro valori  $\epsilon$  si allontanano troppo da 1.

# TEST NON PARAMETRICO DI FRIEDMAN

Non possiamo utilizzare ANOVA, utilizziamo il test di Friedman.

Test di Friedman

<b>Gruppo</b>	<b><math>\chi</math>-squared</b>	<b>p-value</b>
<b>1</b>	26.19	$8.196 \times 10^{-5}$
<b>2</b>	10.762	0.05631
<b>3</b>	13.143	0.02208
<b>4</b>	8.2857	0.1412

Sono nel secondo e il quarto gruppo abbiamo un p-value  $> 0.05$

## Analisi post-hoc sul primo gruppo

Test di Wilcoxon con correzione di Benjamini-Hochberg

<b>Anni</b>	<b>2014</b>	<b>2015</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>
<b>2015</b>	0.438	-	-	-	-
<b>2016</b>	0.108	0.043	-	-	-
<b>2017</b>	0.078	0.043	0.438	-	-
<b>2018</b>	0.043	0.043	0.043	0.043	-
<b>2019</b>	0.043	0.043	0.043	0.043	0.043

# ANOVA A MISURE RIPETUTE

I primi due gruppi non rispettano esclusivamente il vincolo della sfericità, verifichiamo se ANOVA conferma i risultati del test di Friedman

ANOVA a misure ripetute

Gruppo	F-value	p-value
<b>Gruppo 1</b>	4.482	0.0878
<b>Gruppo 2</b>	0.229	0.652

In entrambi i casi accettiamo l'ipotesi nulla:

- Nel primo gruppo utilizzando ANOVA accettiamo l'ipotesi nulla per la quale, con Friedman, vi erano fortissime evidenze contrarie
- Nel secondo gruppo non abbiamo evidenza contro l'ipotesi nulla, come succedeva con Friedman

Tuttavia ANOVA è in questo caso molto meno significativo rispetto al test di Friedman!







# CONCLUSIONI



Nell'analisi presentata:

- In alcuni periodi e gruppi specifici (ad es. paesi ad alto PIL e alto tasso di suicidio per WR), non ci sono differenze significative nella distribuzione annuale del tasso di suicidi.
- Tuttavia, anche confrontando due gruppi con quattro paesi in comune, ma con due differenze, emergono risultati contrastanti (terzo e quarto gruppo).
- Date le diverse cause potenziali del fenomeno, questa analisi fornisce una base per futuri studi che potrebbero integrare questo dataset con altri fattori sociali, economici o geopolitici influenti sulla distribuzione dei tassi di suicidio.

# BIBLIOGRAFIA

-  SHNEIDMAN ES. "THE DEFINITION OF SUICIDE". NEW YORK, NY AND LONDON: JOHN WILEY AND SONS; 1985.
-  GVION Y, APTER A. "SUICIDE AND SUICIDAL BEHAVIOR". PUBLIC HEALTH REVIEWS. 2012;34: EPUB AHEAD OF PRINT.
-  VÄRNIK, P. "SUICIDE IN THE WORLD". INT. J. ENVIRON. RES. PUBLIC HEALTH 2012, 9, 760-771. [HTTPS://DOI.ORG/10.3390/IJERPH9030760](https://doi.org/10.3390/IJERPH9030760)
-  WORLD HEALTH ORGANIZATION. (2023). "SUICIDE RATES". RECUPERATO DA [HTTPS://WWW.WHO.INT/DATA/GHO/DATA/THEMES/MENTAL-HEALTH/SUICIDE-RATES](https://www.who.int/data/gho/data/themes/mental-health/suicide-rates)
-  STATISTICS CANADA. (2023). "AGE-STANDARDIZED RATES". RECUPERATO DA [HTTPS://WWW.STATCAN.GC.CA/EN/DAI/BTD/ASR](https://www.statcan.gc.ca/en/dai/btd/asr)
-  COUNTRYECONOMY.COM. (2023). "GDP - GROSS DOMESTIC PRODUCT 2019". RECUPERATO DA [HTTPS://COUNTRYECONOMY.COM/GDP?YEAR=2019](https://countryeconomy.com/gdp?year=2019)