**Università degli Studi di Salerno**



**Dipartimento di INFORMATICA**

**Progetto di Statistica e Analisi dei Dati**

***Inferenza statistica applicata alla stima di positivi all’alcol test***

***su una popolazione di automobilisti fermati a controllo***

**Docente: Studente:**

***Prof.ssa. Amelia G. Nobile******Ferrara Carmine***

***Matr.05225/00990***

**ANNO ACCADEMICO 2020/2021**

Sommario

[Introduzione 2](#_Toc56438448)

[Problematica in esame e variabile aleatoria 2](#_Toc56438449)

# Introduzione

Con lo studio di statistica descrittiva condotto sulla problematica di analisi del consumo alcolemico in Italia del 2019, è emersa la presenza di valori anomali oppure fortemente elevati per molte regioni d’Italia, in particolare poi, è proprio l’Istituto Superiore di Statistica a ribadire che in Italia, relativamente all’anno 2019, oltre il 66% della popolazione sopra gli 11 anni ha consumato almeno una bevanda alcolica nell’anno.

A fronte di tali stime, è facile evincere quanto sia doveroso, tenere sotto controllo i parametri relativi all’alcolismo al fine di non incorrere in situazioni problematiche. In particolare, questa tematica oltre che da un punto di vista puramente descrittivo, può essere anche studiata e settorializzata in vari contesti della quotidianità nazionale, ed in particolar modo è senz’altro interessante capire come il consumo di alcool può influenzare un determinato fenomeno come l’automobilismo o la medicina.

Considerare però un’indagine statistica su una popolazione molto elevata come ad esempio il mondo degli automobilisti italiani o anche pensare di simulare una realtà così grande, per vedere gli effetti che l’alcool provoca su di essa, è un’operazione molto complessa. Per questo motivo in statistica si preferisce utilizzare quella che viene definita come *Inferenza Statistica*: che si pone l’obbiettivo di analizzare un particolare fenomeno su un determinato campione estratto da una popolazione, e cerca di rapportare i risultati ottenuti dal campione su una popolazione statistica infinita o estremamente grande.

Uno degli obbiettivi principali dell’inferenza statistica, è quello studiare una popolazione descritta da una variabile aleatoria osservabile X, la cui funzione di distribuzione ha una forma nota e riconducibile a standard teorici, ma contiene uno o più parametri non noti che ne caratterizzano l’andamento. Stimare questi parametri non noti, al fine di ottenere una buona analisi, è compito di quella che in statistica inferenziale, prende il nome di stima dei parametri, la quale può essere:

* Puntuale: se per ogni termine non noto si decide di utilizzare un valore fisso;
* Intervallare: se ogni valore non noto, viene invece rilevato da un certo intervallo, detto intervallo di confidenza (da cui appunto si estrae poi un determinato parametro non noto, con un certo grado di confidenza).

Queste stime verranno poi successivamente validate, con un’ulteriore procedura che prende il nome di verifica delle ipotesi, dove appunto secondo alcune congetture e metodologie, si andrà a determinare se il parametro rilevato è adatto o meno per il campione studiato.

# Problematica in esame e variabile aleatoria

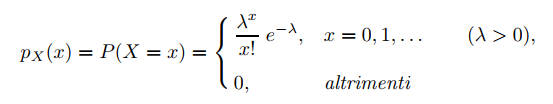
Il tema del consumo di alcol in Italia, così come in altri paesi, è vincolato da stringenti leggi, le quali però molto spesso vengono infrante, mettendo anche in serio pericolo l’incolumità di chi è a stretto contatto con il trasgressore. In particolare, il mondo dell’automobilismo è strettamente vincolato da questa problematica, visto che non pochi automobilisti che vengono sottoposti a controllo risultano avere un tasso alcolemico al di fuori del limite consentito dalla scala giuridica che vieta o consente la guida in relazione ai livelli di i livelli di alcol nel sangue. In particolare, data questa premessa, viene spontaneo chiedersi:

*“Dato una certa popolazione, o un certo campione di automobilisti fermati dalle forze dell’ordine per controlli di routine ed in particolare ad alcol test, quanti di essi risultano avere un tasso alcolemico superiore ai limiti consentiti dalla legge?”.*

Questa problematica, per ogni elemento del campione, risulta essere caratterizzata da un controllo indipendente rispetto agli altri individui, in altri termini, ogni conducente d’auto che viene sottoposto all’alcol test, può risultare in regola o meno rispetto alle norme vigenti, indipendentemente dall’esito di altri conducenti d’auto. In statistica queste tipologie di fenomeni che possono avere soltanto due possibili esiti prendono il nome di prove di Bernulli, ed in particolare quando si considera più di una prova verrebbe da pensare che la variabile più adatta a questa tipologia di fenomeni, sia la variabile discreta Binomiale.

Però come si può effettivamente considerare dalla la problematica in esame, sarebbe poco realistico o molto confusionario utilizzare la distribuzione Binomiale, in quanto pur trattandosi di fenomeni di conteggio, il numero di persone che risulta avere un tasso alcolemico fuori limite, tra le tante entità di un campione che risultano essere sottoposte ad alcol test, sicuramente è un numero molto esiguo. Per effettuare delle stime più corrette rispetto a questa tipologia di dati, è senz’altro più significativo considerare la variabile aleatoria di Poisson, che appunto si presta molto meglio a quelli che in statistica vengono definiti come eventi Rari.

La variabile aleatoria di Poisson è una variabile aleatoria di tipo discreto, che è vincolata dalla seguente funzione di probabilità:



Dove appunto x è appunto il valore che la variabile aleatoria assume nel range di valori che la funzione di probabilità consente (in questo caso da 0 ad infinito), mentre il parametro λ, è un parametro non noto che determina al meglio con che probabilità, la variabile aleatoria X, assume un certo valore x, secondo la legge riportata.

In particolare, è noto che il valore medio che la variabile Aleatoria assume e la varianza tra gli elementi di una data distribuzione Poissoniana (che in probabilità si ricava come momento del secondo ordine – momento del primo ordine al quadrato), risultano essere proprio uguali al parametro non noto λ .



Quest’osservazione, sarà molto importante ai fini dell’indagine da condurre, perché anche successivamente, tramite il parametro λ, si potrà avere una stima quantitativa del valore medio assunto dalla variabile aleatoria.