Significato nomi dei metodi:

pdf (*probability density function*): densità di probabilità
$P(X = x)$

cdf (*cumulative distribution function*): funzione di ripartizione cumulativa
$P(X <= x)$

sf (*survival function*): funzione sopravvivenza
$P(X > x)$

ppf (*percentile point function*): inverso di cdf, quantile
restituisce t tale che $P(X < t) = q$

Distribuzioni discrete

1 Distribuzione binomiale

Densità: binom.pmf(k, n, p)
Funzione di ripartizione: binom.cdf(k, n, p)
Funzione di sopravvivenza: binom.sf(k, n, p)
Parametri:

- k: numero di successi
- n: numero di prove
- p: probabilità di successo

Media: $n \cdot p$
Varianza: $n \cdot p \cdot (1 - p)$

2 Distribuzione ipergeometrica

Densità: hypergeom.pmf$(k, b + r, b, n)$
Funzione di ripartizione: hypergeom.cdf$(k, b + r, b, n)$
Funzione di sopravvivenza: hypergeom.sf$(k, b + r, b, n)$
Parametri:

- k: numero di successi
- b: cardinalità dell'insieme da cui scegliere per avere successi
- r: cardinalità dell'insieme da cui scegliere per avere insuccessi

- n: numero di prove

Media: $n \cdot \frac{b}{b+r}$
Varianza: $n \cdot \frac{b \cdot r}{(b \cdot r)^2} \cdot \frac{b+r-n}{b+r-1}$

3 Distribuzione geometrica

Densità: geom.pmf(T, p)
Funzione di ripartizione: geom.cdf(T, p)
Funzione di sopravvivenza: geom.sf(T, p)
Parametri:

- T: tempo di primo successo
- p: probabilità di successo

Media: $\frac{1-p}{p}$
Varianza: $\frac{1-p}{p^2}$

4 Distribuzione di Poisson

Densità: poisson.pmf(k, λ)
Funzione di ripartizione: poisson.cdf(k, λ)
Funzione di sopravvivenza: poisson.sf(k, λ)
Parametri:

- k: numero di successi
- λ: numero medio di successi attesi

Media: λ
Varianza: λ

5 Distribuzione multinomiale

Densità: multinomial.pmf(k, n, p)
Parametri:

- k: lista di esiti
- n: numero di prove
- p: lista delle probabilità di successo

Distribuzioni continue

6 Distribuzione uniforme

Densità: uniform.pdf$(x, a, b-a)$
Funzione di ripartizione: uniform.cdf$(x, a, b-a)$
Funzione di sopravvivenza: uniform.sf$(x, a, b-a)$
Quantile: uniform.ppf$(q, a, b-a)$
Parametri:

- x: variabile casuale
- a: limite inferiore
- b: limite superiore
- q: ordine del quantile

Media: $\frac{a+b}{2}$
Varianza: $\frac{(b-a)^2}{12}$

7 Distribuzione normale

Densità: norm.pdf(x, μ, σ)
Funzione di ripartizione: norm.cdf(x, μ, σ)
Quantile: norm.ppf(q, μ, σ)
Funzione di sopravvivenza: norm.sf(x, μ, σ)
Parametri:

- x: variabile casuale
- μ: media
- σ: deviazione standard
- q: ordine del quantile

Media: μ
Varianza: σ^2

8 Distribuzione esponenziale

Densità: $\text{expon.pdf}(x, \text{scale} = 1/\lambda)$
Funzione di ripartizione: $\text{expon.cdf}(x, \text{scale} = 1/\lambda)$
Quantile: $\text{expon.ppf}(q, \text{scale} = 1/\lambda)$
Funzione di sopravvivenza: $\text{expon.sf}(x, \text{scale} = 1/\lambda)$
Parametri:

- x: variabile casuale
- scale: scala della distribuzione, inverso del tasso λ (ovvero $1/\lambda$)
- q: ordine del quantile

Media: $1/\lambda$ (o scale)
Varianza: $1/\lambda^2$ (o scale^2)

9 Distribuzione di Student

Densità: $\text{t.pdf}(x, \text{df})$
Funzione di ripartizione: $\text{t.cdf}(x, \text{df})$
Funzione di sopravvivenza: $\text{t.sf}(x, \text{df})$
Quantile: $\text{t.ppf}(q, \text{df})$
Parametri:

- x: variabile casuale
- df: gradi di libertà
- q: ordine del quantile

Media: 0 se $\text{df} > 1$, altrimenti non è definita
Varianza: $\frac{\text{df}}{\text{df}-2}$ se $\text{df} > 2$, altrimenti non è definita

10 Distribuzione Chi-Quadro

Densità: $\text{chi2.pdf}(x, \text{df})$
Funzione di ripartizione: $\text{chi2.cdf}(x, \text{df})$
Quantile: $\text{chi2.ppf}(x, \text{df})$
Funzione di sopravvivenza: $\text{chi2.sf}(x, \text{df})$
Parametri:

- x: variabile casuale
- df: gradi di libertà

Media: df
Varianza: $2 \cdot \text{df}$

11 Distribuzione Normale Multivariata

Densità: multivariate_normal.pdf(x, mean $= None$, cov $= 1$)
Funzione di ripartizione: multivariate_normal.cdf(x, mean $= None$, cov $= 1$)
Parametri:

- x: variabile casuale
- mean: media (se non specificata, assume il valore di default None)
- cov: matrice di covarianza (se non specificata, assume il valore di default 1)

Media: (μ_1, μ_2)
Matrice di covarianza: $\begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$

12 Teorema del limite centrale

Sia Xn una successione di v.a indipendenti e tutte con la stessa legge di media μ e varianza $\sigma^2 > 0$finiti. Allora la variabile normale standardizzata

$$S^* = \frac{\bar{X_{\mathrm{n}}} - \mu}{\sigma} \cdot \sqrt{n}$$

con $\bar{X}_n$ che tende in legge ad una variabile $S \sim N(0,1)$

OSS Nella pratica, per $n \gg 1$, si approssima S_n^* con S di legge $N(0,1)$. Un criterio per tale approssimazione è che $n \geq 30$.

OSS Nel caso particolare della legge B(n,p) un criterio empirico è $n \cdot p >= 5$ e $n \cdot (1-p) >= 5$.

13 Statistica Descrittiva

-Frequenza relativa: $p_i = \frac{f_i}{N}$
-Frequenza relativa cumulata: $F_i = \sum_{k=1}^{i} p_k$
-Media: $\bar{x} = \sum_{k=1}^{n} x_k p_k$
-Varianza Empirica : $\sigma^2 = \sum_{k=1}^{n} (x_k - \bar{x})^2 p_k$
-Quantile della funzione di ripartizione empirica:
$q_\alpha = x_{j-1} + (\alpha - F_{j-1}) * \frac{x_j - x_{j-1}}{F_j - F_{j-1}}$

13.1 FUNZIONI PYTHON PER LA RAPPRESENTAZIONE DEI DATI

Per le rappresentazioni grafiche importare la funzione
import matplotlib.pyplot as plt

-ISTOGRAMMA:
fig, ax = plt.subplot(1,1)
plt.bar(x,y)
plt.show()

parametri:
-x: valori o classi di valori
-y frequenze

-GRAFICO A TORTA:
fig,ax = plt.subplot(1,1)
plt.pie(x,labels="classe di valore",auotpct="formato")
plt.show()

parametri:
-x: frequenze
-autopct="%1.1f%%" per mostrare frequenze in percentuale

-BOX-PLOT:
fig,ax = plt.subplot(1,1)
plt.boxplot(x)
plt.show()

parametri:
-x: frequenze

-GRAFICO QUANTILE-QUANTILE:
per vedere se certi dati seguono una determinata distribuzione
from scipy.stats import probplot
fig,ax = plt.subplot(1,1)
probplot(x,dist="distribuzione da controllare",plot=ax)
plt.show()

parametri:
-x: frequenze
-dist la distribuzione che voglio vedere se i dati vi si adattano

14 STATISTICA INFERENZIALE

-Stimatore media teorica: X_n = media campionaria

-Stimatore varianza teorica con media μ nota

$\sigma^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - u)^2$

-Stimatore varianza teorica con media μ non nota

$\sigma^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{X}_n)^2$

-Intervallo di confidenza per la media con varianza $\sigma^2 nota$

$[\bar{X}_n - \frac{\sigma}{\sqrt{n}}\phi_{1-\frac{\alpha}{2}}, \bar{X}_n + \frac{\sigma}{\sqrt{n}}\phi_{1-\frac{\alpha}{2}}]$
α indice di affidabilità

-Intervallo di confidenza per la media con varianza σ^2 non nota

$[\bar{X}_n - \frac{S}{\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-1), \bar{X}_n + \frac{S}{\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-1)]$

α indice di affidabilità
S stimatore varianza teorica, n dimensione campionaria
n diemnsione camionaria,gradi di libertà della t di stundent
-Intervallo di confidenza per la varianza
$[\frac{S^2(n-1)}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}, \frac{S^2(n-1)}{\chi^2_{\frac{\alpha}{2}}(n-1)}]$

unilatero a destra: $[\frac{S^2(n-1)}{\chi^2_{\alpha}(n-1)}, +\infty]$

unilatero a sinistra $[-\infty, \frac{S^2(n-1)}{\chi^2_{1-\alpha}(n-1)}]$

- intervallo di confidenza per la proporzione

$[\bar{X}_n - \phi_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}, \bar{X}_n + \phi_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}]$

14.1 FUNZIONI PYTHON PER LA STATISTICA INFERENZIALE

importare la libreria numpy
import numpy as np

-per calcolare la deviazione standard
np.std(S,ddof=1)

-per calcolare la media campionaria
np.mean(X)

-per calcolare quantile di ordine α di una certa distribuzione
libreriadistribuzione.ppf(α)