

Nama : Anindita Khusnul Oktavia
Program : Introduction to Python for Data Science

SUMMARY

INFERENCEAL STATISTICS

Inferential statistics mengambil data sebagai sampel dari populasi yang lebih besar untuk membuat kesimpulan. Tujuan utama dari metode ini adalah untuk menarik kesimpulan dari sampel dan menggeneralisasikannya untuk populasi yang lebih besar.

Probability Distributions

Probability distribution adalah fungsi di bawah teori dan statistik probabilitas - yang memberi kita seberapa besar kemungkinan hasil yang berbeda dalam sebuah eksperimen.

Random Variable

Random variable adalah variabel yang kemungkinan nilainya merupakan hasil numerik dari fenomena acak.

- **Discrete random variable** adalah variabel yang mungkin hanya mengambil sejumlah nilai berbeda yang dapat dihitung dan dengan demikian dapat dikuantifikasi.
- **Continuous random variable** adalah variabel yang mengambil jumlah kemungkinan nilai tak terbatas. Misalnya, kita dapat menentukan variabel acak X sebagai tinggi siswa di kelas.

Uniform Distribution

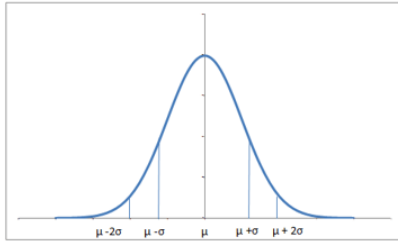
Probability distribution function dari continuous uniform distribution adalah:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b, \\ 0 & \text{for } x < a \text{ or } x > b \end{cases}$$

Setiap interval angka dengan lebar yang sama memiliki probabilitas yang sama untuk diamati, kurva yang menggambarkan distribusi tersebut adalah persegi panjang, dengan tinggi konstan melintasi interval dan 0 tinggi di tempat lain.

Normal Distribution

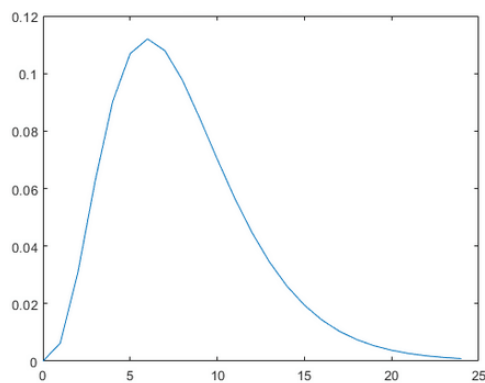
Normal Distribution (Gaussian distribution) memiliki kurva kerapatan berbentuk lonceng yang dijelaskan dengan mean μ dan deviasi standarnya σ . Kurva kepadatan simetris, berpusat di sekitar meannya, dengan penyebarannya ditentukan oleh deviasi standarnya yang menunjukkan bahwa data di dekat mean lebih sering terjadi daripada data yang jauh dari mean. Contoh :



Hampir 68% data berada dalam jarak satu standard deviation dari mean di kedua sisi dan 95% dalam dua standard deviation. Juga perlu disebutkan bahwa distribusi dengan mean 0 dan standard deviation 1 disebut **standard normal distribution**.

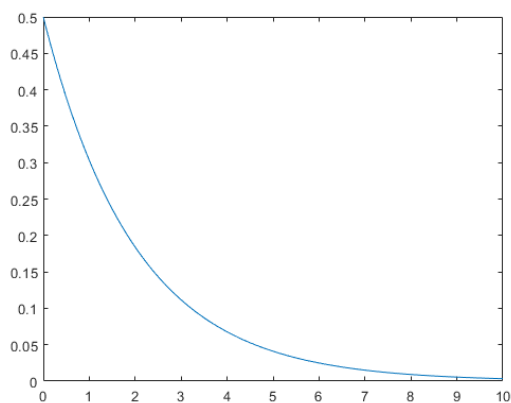
Gamma Distribution

Gamma distribution adalah two-parameter family dari continuous probability distributions. Meskipun jarang digunakan dalam bentuk mentahnya tetapi distribusi lain yang populer digunakan seperti exponential, chi-squared, erlang distributions adalah kasus khusus dari distribusi gamma. Contoh Gamma distribution yaitu :



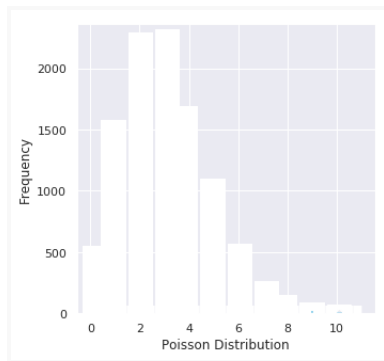
Exponential Distribution

Exponential distribution menggambarkan waktu antara peristiwa dalam Poisson point process, yaitu proses di mana peristiwa terjadi terus menerus dan independen pada tingkat rata-rata yang konstan. Contoh Distribusi eksponensial yaitu :



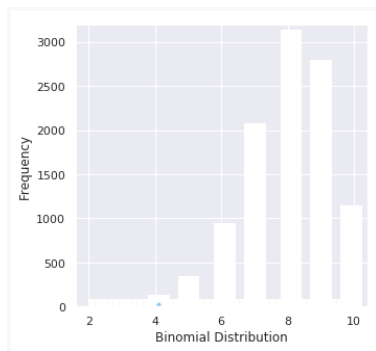
Poisson Distribution

Poisson random variable biasanya digunakan untuk memodelkan berapa kali suatu peristiwa terjadi dalam interval waktu. Contoh Poisson Distribution :



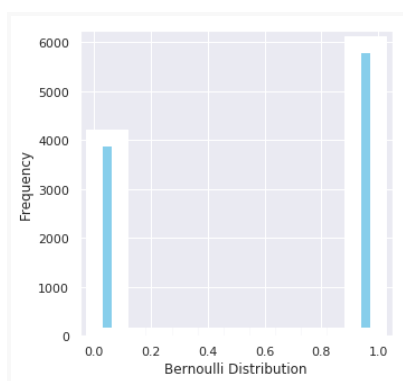
Binomial Distribution

Distribusi di mana hanya dua hasil yang mungkin, seperti sukses atau gagal, untung atau rugi, menang atau kalah dan di mana probabilitas keberhasilan dan kegagalan sama untuk semua percobaan disebut Binomial Distribution. Contoh Binomial Distribution yaitu :



Bernoulli Distribution

Bernoulli distribution hanya memiliki dua kemungkinan hasil, yaitu 1 (berhasil) dan 0 (gagal), dan dalam sekali percobaan, misalnya, lemparan koin. Contoh Bernoulli Distribution yaitu :



Confidence Intervals

Confidence Interval adalah rentang antara dua nilai di mana nilai suatu Sample Mean tepat berada di tengah-tengahnya. Dalam statistik, confidence interval (CI) adalah jenis estimasi yang dihitung dari data statistik yang diamati. Confidence interval mengukur seberapa akurat Mean sebuah sample mewakili (mencakup) nilai Mean Populasi sesungguhnya.

How to Calculate the Confidence Interval

Perhitungan confidence interval melibatkan estimasi terbaik yang diperoleh sample dan margin of error. Rumus confidence interval dan margin of error yaitu:

$$\text{Best Estimate} \pm \text{Margin of Error}$$

$$\text{Margin of Error} = z * \text{Estimated SE}$$

Dengan menggabungkan kedua rumus di atas, dapat diuraikan rumus CI sebagai berikut:

$$\text{Population Proportion or Mean} \pm z - \text{score} * \text{Standard Error}$$

Rumus untuk menghitung standard error dari population proportion adalah:

$$\begin{aligned} &\text{Standard Error For Population Proportion} \\ &= \sqrt{(\text{Population Proportion} * \frac{(1 - \text{Population Proportion})}{\text{Number Of Observations}})} \end{aligned}$$

Rumus untuk menghitung standard error dari sample mean adalah:

$$\text{Standard Error For Mean} = \frac{\text{Standard Deviation}}{\sqrt{\text{NumberOf Observations}}}$$

Untuk implementasinya ada pada file .ipynb