2 - PROBABILIDADE

ÁREAS DA ESTATÍSTICA

1

Estatística Descritiva

(usada no processo de análise)

2

Probabilidade

(usada nos algoritmos de machine learning)

3

Estatística Inferencial

(usada para fazer inferencias sobre uma população/amostra)

PROBABILIDADE

2

- Estudo da aleatoriedade e da incerteza (medida da incerteza).
- A probabilidade é um número que varia de 0 a 1 e que mede a chance de ocorrência de um determinado resultado. Pode ser expressa através de decimais, frações e percentagens.
- Princípio básico do aprendizado humano que é a ideia de experimento.

Experimentos Aleatórios (casuais)

Experimentos Não Aleatórios (determinísticos [resultado já é sabido])

EXPERIMENTO ALEATÓRIO

Experimento Aleatório é o fenômeno que, quando repetido inúmeras vezes em processos semelhantes, possui resultados imprevisíveis.

Experimento é qualquer atividade realizada que pode apresentar diferentes resultados. Um experimento é dito aleatório quando não conseguimos afirmar o resultado que será obtido antes de realizar o experimento. Um experimento é dito equiprovável se todos os possíveis resultados possuem a mesma chance de ocorrer.

Evento: um ou mais resultados de um experimento.

O resultado(s) são um subconjunto do espaço da amostra.

EX:
Experimento: JOGAR UM DADO
Espaço da amostra: 6

TIPOS DE PROBABILIDADE

Probabilidade Clássica

Usada quando sabemos o número de possíveis resultados do evento e podemos caclular a probabilidade do evento com a formula:

$$P(A) = \frac{\textit{Número de possíveis resultados do evento } A}{\textit{Numero total de possíveis resultados dentro do espaço da amostra}}$$

Onde: P(A) é a probabilidade de um evento ocorrer.

Probabilidade Empírica

Envolve conduzirmos um experimento, para observarmos a frequência com que um evento ocorre.

$$P(A) = \frac{\textit{Frequência em que o evento } A \textit{ ocorre}}{\textit{Número total de observações}}$$

Probabilidade Subjetiva

Usamos quando dados ou experimentos não estão disponíveis para calcular a probabilidade.

REGRAS DA PROBABILIDADE

1ª Regra — Se P(A) =1, então podemos garantir que o evento A **ocorrerá**

2ª Regra — Se P(A) =0, então podemos garantir que o evento A **não ocorrerá**

3ª Regra — A probabilidade de qualquer evento sempre será entre **0 e 1.**

4ª Regra — A soma de todas as probabilidade para um evento simples, em um espaço de amostra, será **igual a 1.**

5ª Regra — O complemento do envento A é definido como todos os resultados em um espaço de amostra, que **não** fazem parte do envento A. Ou seja:

P(A) = 1 – P(A’), onde P(A’) é o complemento do evento A.

EVENTO E ESPAÇO AMOSTRAL

Espaço Amostral é o conjunto de possíveis resultados de um experimento ou fenônemo aleatório, representado por S.

Evento é qualquer subconjunto do espaço amostral S de um evento aleatório.

Eventos Complementares: p (sucesso) q (insucesso) para um mesmo evento existe sempre a relação:

$p + q = 1 \rightarrow q = 1 - p$

Eventos Independentes: quando a realização ou não de um evento não afeta a probabilidade da realização do outro evento.

Para eventos se realizarem simultaneamente: $p = p1 \times p2$

PROBABILIDADE CONCIDIONAL :
ALGORITMO NAÏVE BAYES (CLASSIFICAÇÃO)

Probabilidade Conjunta: Probabilidade de dois eventos ocorrerem juntos.

Independência: Regra Geral de Multiplicação para eventos compostos.

$$P(A\ e\ B) = P(A) * P(B|A)$$

Probabilidade Condicional: condição previa

P(B | A) pronunciamos “a probabilidade de B dado A”. Onde P(A) ≠ 0:

$$P(B|A) = \frac{P\ (A\ and\ B)}{P(A)}$$

RESUMO – PROBABILIDADE CONDICIONAL E INDEPÊNDENCIA

Ou (Or)	Em Geral	P(A or B) = P(A) + P(B) - P(A and B)
	Se eventos são mutuamente exclusivos	P(A or B) = P(A) + P(B)
E (And)	Em Geral	P(A and B) = P(A) * P(B | A) = P(A | B) * P(B)
	Se eventos são independentes	P(A and B) = P(A) * P(B)

TABELA DE CONTINGENCIA

Tabela 2

Tabela de Contingência R x C no h-ésimo Estrato

Níveis do Fator	Categorias da variável de resposta						
	1	2	.	j	.	C	Total
1	n_{h11}	n_{h12}	.	n_{h1j}	.	n_{h1C}	$N_{h1\cdot}$
2	n_{h21}	n_{h22}	.	n_{h2j}	.	n_{h2C}	$N_{h2\cdot}$
$\vdots$	$\vdots$	$\vdots$	.	$\vdots$	.	$\vdots$	$\vdots$
i	n_{hi1}	n_{hi2}	.	n_{hij}	.	n_{hiC}	$N_{hi\cdot}$
$\vdots$	$\vdots$	$\vdots$	.	$\vdots$	.	$\vdots$	$\vdots$
R	n_{hR1}	n_{hR2}	.	n_{hRj}	.	n_{hRC}	$N_{hR\cdot}$
Total	$N_{h\cdot 1}$	$N_{h\cdot 2}$	.	$N_{h\cdot j}$	.	$N_{h\cdot C}$	$N_{h\cdot\cdot}$

DISTRIBUIÇÃO DE PROBABILIDADE – VARIÁVEL ALEATÓRIA

- Uma **Distribuição de Probabilidade** descreve a **chance** que uma variável (discreta ou continua) pode assumer ao longo de um espaço de valores.
- Uma **Distribuição de Probabilidade** é um modelo matemático que relaciona um certo valor da variável de estudo com a probabilidade de ocorência.
- O conjunto de todos os possíveis resultados de um experimento aleatório é denominado espaço amostral.

- O resultado de um experimento de probabilidade geralmente é uma contagem ou uma medida. Quando isso ocorre, o resultado é chamado de **variável aleatória** **(discrete ou continua)**.

TIPOS DE DISTRIBUIÇÃO DE PROBABILIDADE

- Distribuição Discreta
  - Binomial
  - Poisson
  - Bernoulli
  - Geométrica
  - Hipergeométrica
- Distribuição Contínua
  - Normal
  - Uniforme
  - Exponencial
  - Gama
  - t de Student

DISTRIBUIÇÃO DE PROBABILIDADE DISCRETAS

Binomial

- É utilizada para descrever cenários em que os resultados de uma variável aleatória podem ser agrupados em duas categorias. As categorias dessa distrib são classidicadas como: **Sucesso (p)** ou **Falha (q)** **(p = 1 – q)**
- Media : $\mu = n * p$
- Variância: $\sigma^2 = (n.p).(1 - p)$

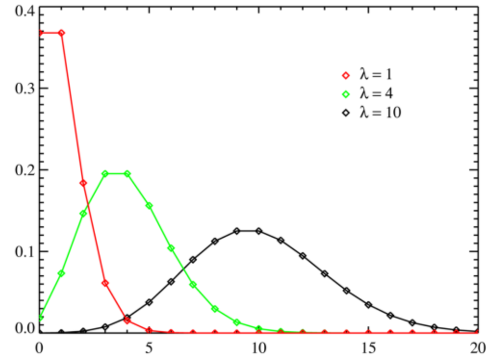
Onde *n = número de tentativas, p = probabilidade de sucesso.*

Poisson

- É usada para descrever cenários onde existe a probabilidade de ocorrencia do envento em um intervalo contínuo.
- A Distribuição Poisson é caracterizada pelo parâmentro único chamado λ (lambda), que representa a taxa media de ocorrência por unidade de medida.

Nas Distribuições Binomial e Poisson os eventos são independentes uns dos outros. Desta forma, a probabilidade de sucesso ou de número de ocorrências, se mantém constante.

$$P(x) = \frac{e^2.\lambda^x}{x!}$$



DISTRIBUIÇÃO DE PROBABILIDADE DISCRETAS

Hipergeométrica

- A **Distribuição Hipergeométrica** é uma distribuição de probabilidade discreta que descreve o número de sucessos numa sequência de **n** extrações de uma população finite, ou seja, sem resposição.
- Quando a amostragem é sem reposição, a probabilidade de **sucesso muda** durante o processo

de amostragem, isso viola os requisitos para uma distribuição de probabilidade binomial. Nesse caso usamos a **Distribuição Hipergeométrica.**

$$P(x) = \frac{{}_{N-R}C_{n-x} \cdot {}_{R}C_{x}}{{}_{N}C_{n}}$$

onde:

N = Tamanho da população
R = O número de sucessos da população
n = Tamanho da Amostra
x = Número de sucessos da amostra

$$\mu = \frac{nR}{N}$$

onde:

N = Tamanho da população
R = O número de sucessos da população
n = Tamanho da Amostra

$$\sigma = \sqrt{\frac{nR(N-R)}{N^2}}\sqrt{\frac{N-n}{N-1}}$$

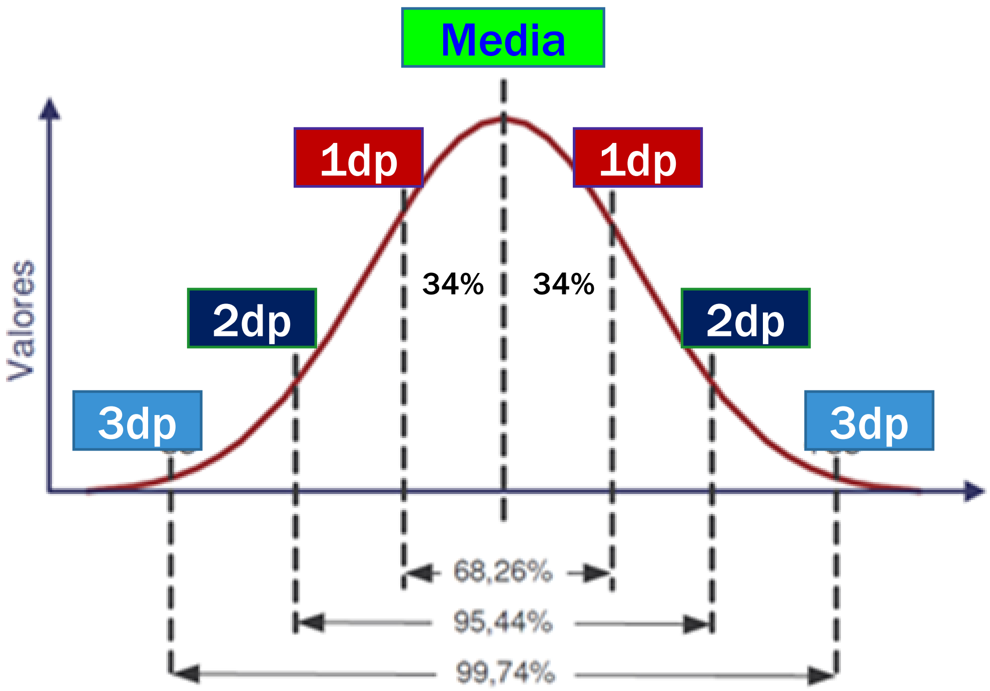
onde:

N = Tamanho da população
R = O número de sucessos da população
n = Tamanho da Amostra

DISTRIBUIÇÃO DE PROBABILIDADE CONTÍNUAS

Normal

- Distribuição Normal (Gaussiana) é uma curva simétrica em torno do seu ponto médio, apresentando assim seu famoso formato de sino (média = 0 e desvio padrão = 1).



Distribuição Normal é útil quando os dados tendem a estar próximos ao centro da distribuição (próximos da média) e quando valores extremos (outliers) são muito raros.

Modelos normais são definidos por dois parâmetros:

μ = média

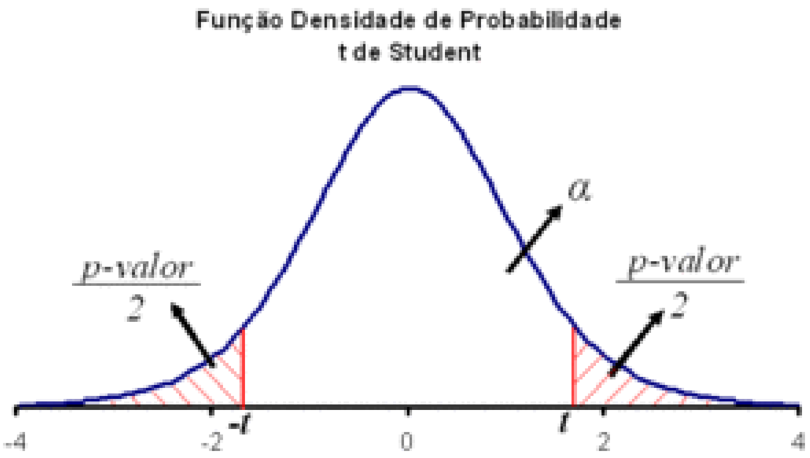
σ = desvio padrão

Em Machine learning temos que padronizar os dados: há um Modelo Normal diferente para cada combinação de μ e σ , mas se padronizarmos nossos dados primeiro, criando z-scores e subtraíndo da média para fazer μ = 0 e dividindo pelo desvio padrão para fazer o σ = 1, então precisaremos apenas do modelo com a média 0 e o desvio padrao 1. Chamamos isso de Modelo Normal Padrão ou Distribuição Normal Padrão (Standard Normal Distribution).

DISTRIBUIÇÃO DE PROBABILIDADE CONTÍNUAS

t de Student

- Uma das principais distribuições da probabilidade, com inúmeras aplicações em inferência estatística



- **Distribuição Uniforme** – um número finito de resultados com chances iguais de acontecer. Usada quando assuimos intervalos iguais da variável aleatória que tem a mesma probabilidade.

Uniforme

- **Distribuição Exponencial** – é usada paradescrever os dados quando valores mais baixos tendem a dominar a distribuição e quando valores muito altos não ocorrem com frequência.
- Na distribuição Exponencial a variável aleatória é definida como o tempo entre as duas ocorrências, sendo a média de tempo as ocorrencias de 1/λ.

Exponencial

DICA – DISTRIBUIÇÃO NORMAL

Obviamente, não devemos usar um modelo Normal para todos os conjuntos de dados. Se o histograma não for em forma de sino, as pontuações (scores) z não serão bem modeladas pelo modelo Normal. E a padronização não ajuda, porque a padronização não altera a forma da distribuição. Portanto, sempre verifique o histograma dos dados antes de usar o modelo Normal.

Padronizar os dados antes de apresentar à um modelo de ML:

OBS: O fato de padronizar os dados não significa que você transformou a distribuição em Normal (pois um Modelo Normal pode ter qualquer media e desvio padrão), para ter uma distribuição normal é necessário ter media 0 e desvio padrao 1.

- *Aplicar o z-score*
- *Fazer as devidas conversões*
- *No Modelo Normal aplicar as conversoes e isso irá resultar em uma distribuição normal.*

COMO DETERMINER SE A DISTRIBUIÇÃO É NORMAL?

Para determiner se uma determinada variável aleatória segue uma distribuição normal, basta verificar se essa segue a função densidade de probabilidade, dada por:

$$f(x) = \frac{e^{\frac{-(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma}}$$

Onde μ é a média e σ^2 é a variância de x.

A notação $N(\mu,\sigma^2)$ é usada para representar tal distribuição. Para calcularmos então a probabilidade de um resultado, basta integrar a função f(x) em relação a x, com os limites de integração representando a faixa de valores que se quer obter a probabilidade. Vale notar que a integral da função densidade de probabilidade normal não possui solução analítica. Sendo assim, seu cálculo deve ser realizado através de um método numérico.

Continua

$$f(x) = \frac{e^{\frac{-(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma}}$$

Onde μ é a média e σ^2 é a variância de x.

Para sanar tal dificuldade a função pode ser padronizada com a substituição dos parâmetros por $\mu = 0$ e $\sigma^2 = 1$. Essa abordagem é dada pela definição de uma nova variável aleatória Z, chamada de variável aleatória normal padronizada. Se x for uma variável aleatória normal com média $E(x) = \mu$ e variância $V(x) = \sigma^2$, a variável aleatória $Z = (x-\mu)/\sigma$ será uma variável aleatória normal, com $E(Z) = 0$ e $V(Z) = 1$.

Ou seja, Z é uma variável aleatória normal padrão.

Dessa forma, é possível obter a área sob a curva da normal padrão de forma analítica, e então obter a área entre dois pontos sob a curva, diretamente com o uso de uma tabela de conversão, e essa área representa uma probabilidade.

RESUMINDO

Na caracterização das distribuições de probabilidade é de grande importância a utilização de medidas que indiquem aspectos relevantes da distribuição, como medidas de posição (media, mediana e moda), medidas de dispersão (varância e desvio padrão) e medidas de assimetria e curtose.

1

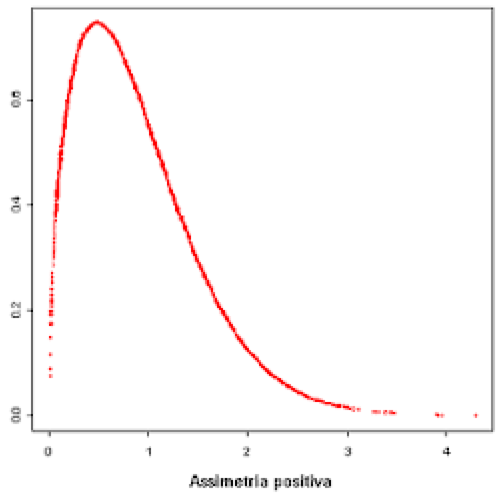
TEOREMA DO LIMITE CENTRAL

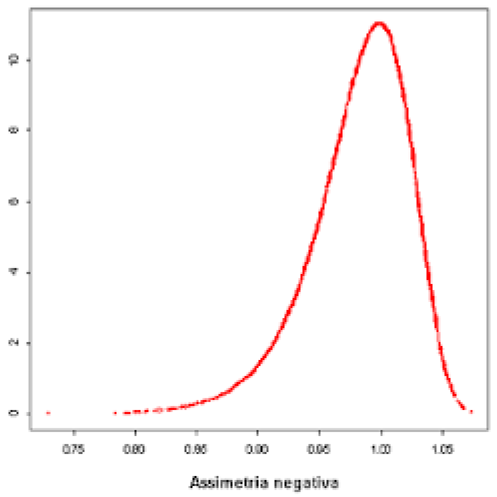
- Esse teorema afirma que quando a amostra aumenta, a distribuição amostral da media tende a aproximar-se cade vez mais de uma distribuição normal.
- O **Teorema do Limite Central** é fundamental para a Estatística uma vez que diversos procedimentos estatísticos comuns requerem que os dados sejam aproximadamente **normais** e o TLC permite aplicar esses procedimentos úteis a popuções que são fortemente **não-normais**.
- Esse teorema possibilita medir o quanto sua media amostral irá variar, sem ter que pegar outra media amostral para fazer a comparação. Ou seja, permite-nos conduzir alguns procedimentos de inferência se ter qualquer conhecimento de distribuição da população.
- **Ele diz que sua media amostral tem uma Distribuição Normal, independente da aparência da distribuição dos dados originais.**

TEOREMA DO LIMITE CENTRAL

- **Muitos procedimentos pressupõem que uma Distribuição Normal é uma Distribuição Simétrica.**
- Assimetria indica variação no formato de distribuição.
- Assimetria positiva: Implica emuma concentração maior de valores menores, e o gráfico possuirá uma cauda mais longa à direita.
- Assimetria negativa: Implica em uma concetração de valores maiores, e o gráfico possuirá uma cauda maior à esquerda.

Def de outlier: é uma valor que está além de 3 desvios-padrão(+-) da media. (gráfico da destri. Normal)



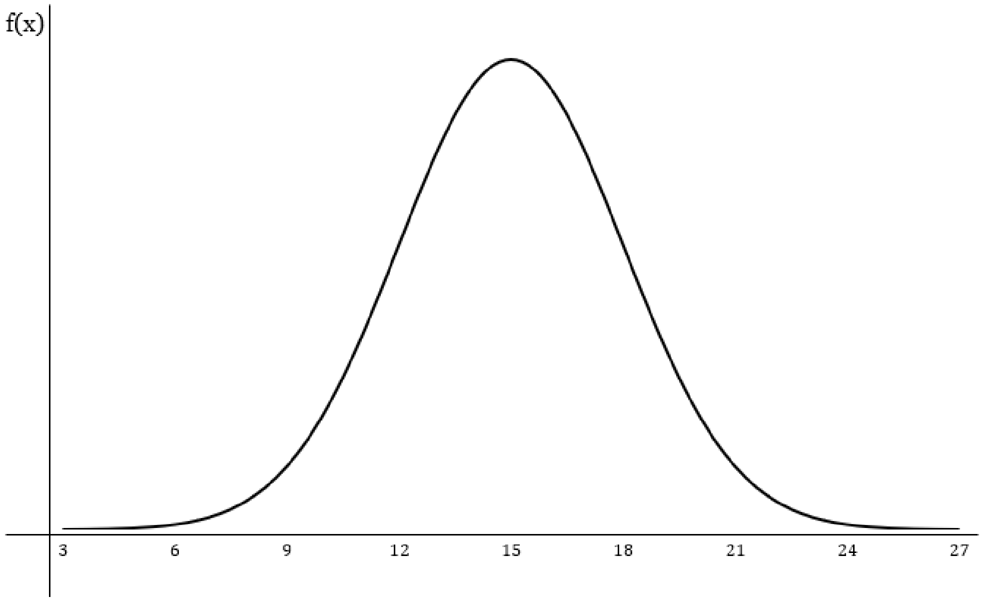


TEOREMA DO LIMITE CENTRAL

A área abaixo da curva Normal representa 100% de probabilidade associada a uma variável.

A probabilidade de uma variável aleatória tomar um valor entre dois pontos quaisquer é igual à área compreendida entre esses dois pontos.

Normal Padrão



End!