Métodos Quantitativos para a Gestão Ambiental

Probabilidades e Distribuições Estatísticas – Parte 1

(4/13)

Luiz Carlos Estraviz Rodriguez

Distribuição de probabilidades

Contexto

O porquê desta aula

Ao desenvolvermos uma “caixa de ferramentas” para o gestor ambiental
começamos pela busca de métodos para a análise exploratória de dados.

Vimos que tabelas e histogramas de frequência revelam importantes aspectos de um conjunto de dados.

As tabelas ou histogramas de frequência expressam a ***distribuição*** de uma determinada variável ou atributos de uma população.

Assim, para uma série histórica da cidade de Piracicaba, vimos como os dias se apresentaram distribuídos entre classes de temperatura, de umidade relativa e de precipitação.

Continuaremos a análise exploratória aprofundando-nos no conceito de ***distribuição***.

Objetivo desta aula

Aula Anterior

Resumimos dados com base na frequência de ocorrência.
E essas frequências foram representadas através de tabelas e histogramas.

Se os dados se referirem a valores de uma variável estocástica (ou seja, de comportamento aleatório), e se o número de observações for suficientemente grande, é possível usar essas frequências para inferirmos sobre o tipo de distribuição de probabilidades que essa variável apresenta.

objetivo da

Aula de Hoje

conhecer e aplicar o conceito de distribuição de probabilidades

Definindo distribuição de probabilidades

Em estatística, uma distribuição de probabilidade descreve a chance de uma determinada variável assumir um determinado valor ao longo de um espaço de valores.

A distribuição de probabilidade é uma função:

- cujo *domínio* são os valores da variável e
- cuja *imagem* são as probabilidades da variável assumir cada valor do domínio.

O conjunto imagem de valores dessa função se restringe ao intervalo entre 0 e 1.

Distribuições para variáveis discretas e contínuas

Uma distribuição de probabilidade pode ser expressa por uma função quando o domínio assumir valores **discretos** (variáveis qualitativas, ordinais ou nominais)

Exemplo: jogo de dados (domínio é formado pelo conjunto finito de valores {1,2,3,4,5,6} e a imagem é a probabilidade de sair um desses valores)

ou

contínuos (variáveis podem assumir qualquer valor dentro de um intervalo)

Exemplo: altura de pessoa (domínio é o infinito conjunto de valores num intervalo razoável de alturas, e imagem é a probabilidade de ocorrer um desses valores)

Recorreremos agora a funções matemáticas para expressar algumas das mais importantes distribuições de probabilidade

Definindo função densidade de probabilidade (*fdp*)

Para cada variável aleatória discreta ou contínua *x* corresponde uma função densidade de probabilidade (***fdp***) que lhes atribui medidas de probabilidade.

A probabilidade de acontecer todo o espaço possível de valores é 1.

E função cumulativa de probabilidades (***fcp***) P{x ≤ X} pode ser definida como:

$$P(X) = \sum_{x=a}^{X} p(x) \qquad \text{se discreta,}$$

$$\text{e } F(X) = \int_a^X f(x)dx \qquad \text{se contínua.}$$

Distribuição de probabilidades - variáveis discretas

Distribuições de probabilidade para variáveis discretas

probabilidade da variável aleatória X assumir um certo valor x:

P(X=x)

A soma de todas as possibilidade que o x pode assumir terá o valor 1 (100%).

As funções de distribuição de probabilidade para variáveis discretas com aplicações interessantes na área de gestão são: *Binomial*, *Poisson* e *Exponencial Negativa*.

Distribuição de probabilidades - variáveis discretas

Distribuição Binomial

Propriedades:

1. As observações são obtidas em *n* ensaios (*ocorrências*) idênticos
2. Em cada ensaio observa-se apenas um dentre dois possíveis valores (sucesso / falha)
3. A probabilidade de sucesso em cada ensaio é *p,* e *p* permanece o mesmo entre ensaios
4. Os ensaios são independentes, ou seja, o resultado de um ensaio não afeta nenhum outro ensaio
5. A variável randômica (ou estocástica) *x* é o número de sucessos observados em *n* ensaios

Distribuição de probabilidades - variáveis discretas

Exemplos de distribuição Binomial

Muitas ‘populações’ de *0s* e *1s* são de interesse para engenheiros, cientistas e empresários:

- A resposta à pergunta “Você é a favor do desenvolvimento da energia nuclear, sim ou não?”
- Experimentação para determinar o efeito de uma nova droga em cobaias
- Processos de monitoramento da qualidade, para determinar a fração da produção com ou sem defeitos

Distribuição de probabilidades - variáveis discretas

Distribuição Binomial

Um processo produz lotes com n itens. A fração p com itens defeituosos por lote é estimada a partir de dados históricos. A questão é determinar a função densidade de probabilidade (fdp) do número de defeitos por lote.

Quantas combinações diferentes são possíveis ao considerar a existência de x defeitos por lote de n itens?

$$C_x^n = \frac{n!}{x!(n-x)!}$$

A probabilidade de obter cada uma dessas combinações é $p^x (1-p)^{n-x}$.

Pela lei da adição de probabilidades, deduz-se que:

$P\{x=k\} = C^n_k \; p^k (1-p)^{n-k} \quad k = 1, 2, ..., n$

Essa é a distribuição binomial com parâmetros n e p. A média e variância são:

$E\{x\} = n\,p$

$var\,\{x\} = n\,p\,(1-p)$

Distribuição de probabilidades - variáveis discretas

Distribuição Poisson

Sugerida em 1837 por S. D. Poisson, esta distribuição tem as seguintes propriedades:

1. Os eventos acontecem um de cada veza, ou seja, dois ou mais eventos não acontecem precisamente no mesmo momento e local (ou espaço)
2. A ocorrência de um evento em um certo tempo, região ou espaço é independente da ocorrência do evento em uma sobreposição desse período, região ou espaço
3. O número esperado de eventos em um período ou região λ é o mesmo que aquele esperado para qualquer outro período ou região

Distribuição de probabilidades - variáveis discretas

Exemplos de distribuição Poisson

Esta é uma distribuição que caracteriza bem processos que formam filas, onde o comprimento da fila depende do número de chegadas

- em um balcão de atendimento (bancos, serviços públicos, posto de saúde etc.)
- em um caixa de pedágio ou estacionamento
- em um posto de inspeção

Distribuição de probabilidades - variáveis discretas

Distribuição Poisson

Clientes chegam de forma totalmente ao acaso (randomicamente), ou seja, é impossível prever quando alguém chegará. A fdp que descreve o número desse tipo de evento (chegadas) durante um período de tempo segue a distribuição Poisson. Seja x o número de eventos (p.ex.: chegadas) num determinado período de tempo (p.ex.: minuto ou hora), a fdp Poisson será definida da seguinte forma:

$$P\{x=k\} = \frac{\lambda^k e^{-\lambda}}{k!} \qquad k = 1, 2, ..., n$$

sendo a média e a variância definidas da seguinte forma:

$$E\{x\} = \lambda$$

$$var\ \{x\} = \lambda$$

Intuitivamente, $E\{x\} = \lambda$ deve representar o número médio de eventos que ocorrem por unidade de tempo. Essencialmente, o parâmetro λ é definido como uma taxa (número por unidade de tempo) à qual o evento ocorre.

<u>Esta distribuição é fundamental para a teoria de filas</u>.

Distribuição de probabilidades - variáveis discretas

Distribuição Exponencial Negativa

Se o número de chegadas a um centro de serviços durante um período específico ocorre de acordo com a distribuição Poisson, então, automaticamente, a distribuição dos intervalos entre chegadas sucessivas segue uma distribuição exponencial negativa (ou, simplesmente, exponencial). Especificamente, se λ é a taxa à qual o evento com distribuição Poisson ocorre, então a distribuição do tempo, *x*, entre chegadas sucessivas é dado por:

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0$$

A média e variância são:

$$E\{x\} = \frac{1}{\lambda}$$

$$var\,\{x\} = \frac{1}{\lambda^2}$$

A média E{x} é consistente com a definição de λ. Se a taxa à qual o evento ocorre, então *1/λ* é o intervalo médio entre eventos sucessivos.

teste de assimilação – primeira parte