Aulas 2-4: Distribuições de Probabilidade e Testes de Hipótese

Análise Quantitativa de Dados Ambientais

Thiago S. F. Silva - tsfsilva@rc.unesp.br

Programa de Pós Graduação em Geografia - IGCE/UNESP

April 10, 2019

Outline

Distribuição de Poisson

Poisson: A Binomial modela o número de sucessos esperados com um numero fixo de realizações. A distribuição Poisson modela a ocorrência de sucessos em situações onde o número de realizações é infinito. O exemplo mais comum são dados de contagem de indivíduos em parcelas, ou ao longo de um intervalo de tempo.

Distribuição Poisson: $X \sim Pois(\lambda)$
Parâmetros: λ $(\lambda > 0)$
Suporte: $x = k, k \in \{0, \dots, n\}$
p.m.f:

$$f(x) = \frac{lambda^k}{k!} e^{-\lambda}$$

$E[X] = \lambda$
$Var[X] = \lambda$

Ex.: Se em média eu observo 50 indivíduos por parcela, qual a probabilidade de eu observar uma parcela com 100 indivíduos?

$$f(x) = \frac{lambda^k}{k!} e^{-\lambda}$$

$$P(x = 100) = \frac{50^{100}}{100!} \times e^{-50}$$

$$P(x = 100) = 845272575844 \times 1.92875 \times 10^{-22}$$

$$P(x = 100) = 1.630319 \times 10^{-10}$$

```
> dpois(100,50)
[1] 1.630319e-10
```

Distribuições Contínuas

- Até agora falamos de V.A. discretas
- Mas e se os dados que queremos modelar são contínuos?
- **Exemplo:** Qual a probabilidade da temperatura máxima de hoje ser 32°C?

Distribuições Contínuas

- Uma V.A. contínua pode assumir infinitos valores

- Se assumimos que: $P \approx F = \frac{n_i}{N}$

- Qual dos dois resultados tem probabilidade maior?

$$P(\text{temperatura máxima de hoje}) = 32^\circ\text{C?}$$

ou

$$P(\text{temperatura máxima de hoje}) = 32.354321^\circ\text{C?}$$

Distribuições Contínuas

Qual dos dois resultados tem probabilidade maior?

P(temperatura máxima de hoje) $= 32°$C?
ou
P(temperatura máxima de hoje) $= 32.354321°$C?

Os dois tem a mesma probabilidade, que é **zero**.

$$P(n_i) \approx F = \frac{n_i}{N} = \frac{1}{\inf} = 0$$

$$\lim_{N \to \inf} P(x) = 0$$

Distribuições Contínuas

Para V.A. contínuas, ao invés de massas de probabilidade, falamos de **densidades de probabilidade**, dentro de um intervalo de valores.

As distribuições de probabilidade contínuas tem, desta maneira, **funções de densidade de probabilidade (f.d.p ou <u>p.d.f</u>)**

Qual a probabilidade da temperatura máxima de hoje estar entre 32 e 33°C?
Qual a probabilidade da temperatura máxima de hoje ser maior que 32°C?

Distribuições Contínuas

Qual a distribuição contínua mais utilizada?

Distribuições Contínuas

Qual a distribuição contínua mais utilizada?

Distribuição Normal (Gaussiana): $X \sim N(\mu, \sigma)$
Parâmetros: μ ($\mu \in \mathbb{R}$)); σ ($\sigma > 0$)
Suporte: $X \in \mathbb{R}$
p.d.f:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

$E(X) = \mu$
$Var(X) = \sigma^2$ (Muitas vezes usamos o desvio padrão: $\sqrt{\sigma^2} = \sigma$)
Exemplo: Qual a probabilidade de observamos uma temperatura entre 30°C e 35°C, se a média histórica é $\mu = 30$, e o desvio padrão é $\sigma = 5$?

Distribuição Normal - Exemplo

Distribuição Normal - Exemplo

```
> # No R
> media <- 30
> desvio <- 5
> tmin <- 30
> tmax <- 35
> # A maneira mais fácil de calcular uma probabilidade é de
> #O comando "pnorm" faz isso:
>
> p.tmin <- pnorm(tmin,mean=media,sd=desvio)
> p.tmin
[1] 0.5
>
```

Distribuição Normal - Exemplo

Distribuição Normal - Exemplo

```
> # No R
> media <- 30
> desvio <- 5
> tmin <- 30
> tmax <- 35
> # calculamos também a probabilidade cumulativa de x <= 35
>
> p.tmax <- pnorm(tmax,mean=media,sd=desvio)
> p.tmax
[1] 0.8413447
>
```

Distribuição Normal - Exemplo

Distribuição Normal - Exemplo

```
> # Sabendo as duas probabilidades cumulativas, é só subtra
>
> p.tmax - p.tmin
[1] 0.3413447
>
```

Distribuição Normal - Exemplo

Distribuições Contínuas

- **Normal**: $X \sim N(\mu, \sigma)$
- **Gamma**: $X \sim Gamma(s, a)$
- **Exponencial**: $X \sim Exp(\lambda)$
- **Beta**: $X \sim Beta(a, b)$
- **Lognormal**: $X \sim logN(\mu, \sigma)$
- **Qui-quadrado**: $X \sim \chi^2(df)$