Formulário de Estatística

Função probabilidade (Espaço amostral Ω)

$$0 \leq p(A) \leq 1 \quad , \quad \forall A \in \Omega$$

$$P(\Omega) = 1 \quad ; \quad P(\emptyset) = 0$$

Probabilidade de A ou B ocorrerem

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Probabilidade condicional

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Independência. A e B são independentes se

$$P(A/B) = P(A)$$

$$P(A \cap B) = P(A) \cdot P(B)$$

Distribuição de probabilidade $f(x) \geq 0$

Variável discreta Xi: $\sum_i P(X_i) = 1$

Variável contínua x: $\int_{-\infty}^{\infty} f(x)dx = 1$

$$P(x \in [a, b]) = \int_a^b f(x)dx$$

Função repartição ou função de distribuição cumulativa $F(a)$

$$F(a) = \int_{-\infty}^{a} f(x)dx$$

Média $\mu = E(x)$

$$\mu = \sum_i X_i \cdot P(X_i)$$

$$\mu = \int_{-\infty}^{\infty} x \cdot f(x) \cdot dx$$

Variância

$$\sigma^2 = \sum_i (X_i - \mu)^2 \cdot P(X_i)$$

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) \cdot dx$$

Desvio padrão: $\sigma = \sqrt{\sigma^2}$

Propriedade da média

k constante: $E(k) = k$

$$E(X \pm Y) = E(X) \pm E(Y)$$

α constante: $E(\alpha \cdot X) = \alpha \cdot E(X)$

X e Y independentes:

$$E(X \cdot Y) = E(X) \cdot E(Y)$$

Propriedades da variância $var = \sigma^2$

$$var(X) \geq 0$$

k constante: $var(k) = 0$

α constante: $var(\alpha \cdot X) = \alpha^2 \cdot var(X)$

X e Y independentes:

$$var(X \pm Y) = var(X) + var(Y)$$

$$var(X) = E(X^2) + E(X)^2$$

Distribuição de Bernoulli

$$X = \begin{cases} 1 & se\ sucesso \\ 0 & se\ fracasso \end{cases} \qquad \begin{cases} P(1) = p \\ P(0) = 1 - p \end{cases}$$

$$\begin{cases} E(X_i) = p \\ \sigma^2(X_i) = p(1 - p) \end{cases}$$

Distribuição Binomial

· **n experim. de Bernoulli independ. X_i**

$$\begin{cases} E(X) = n \cdot p \\ \sigma^2(X) = n \cdot p(1 - p) \end{cases}$$

Probabalidade:

$$P_p(n, k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}$$

Distribuição de Poisson

· **Não há ocorrências simultâneas**

· $\boldsymbol{\lambda = ocorrências/unidade}$

· $\boldsymbol{X_t = nº\ ocorrências/(intervalo\ t)}$

$$P(X_t = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

Distribuição uniforme

$$f(x) = \begin{cases} \dfrac{1}{(b - a)} & se\ x \in [a, b] \\ 0 & caso\ contrário \end{cases}$$

Distribuição exponencial (parâmetro λ)

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & se\ x \geq 0 \\ 0 & caso\ contrário \end{cases}$$

$$F(a) = 1 - e^{-\lambda a}$$

$$E(x) = 1/\lambda \quad ; \quad \sigma^2 = 1/\lambda^2$$

Distribuição normal (parâmetros μ e σ)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot exp\left[-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right]$$

$$X = N(\mu, \sigma^2) \ \rightarrow \ Z = N(0, 1)$$

$$P(a \leq X \leq b) = P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right)$$

__Teorema Central do Limite__

Sejam $X_1, X_2, \dots, X_n$ variáveis independ. com média μ e variância σ^2. Então

$$\bar{X} = \frac{X_1, X_2, \dots, X_n}{n}$$

Tende a uma distribuição normal

$$N(\mu, \sigma^2/n)$$

Aproximação da binomial pela normal:

$$X = X_1 + X_2 + \cdots + X_n = B(n, p)$$

Se $n > 30$, tem-se:

$$X \cong N\big(n \cdot p\,; n \cdot p(1-p)\big)$$

__Estatística descritiva__

Média amostral:

$$\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}$$

Para dados agrupados:

$$\bar{X} = \frac{\sum_{i=1}^{j} f_i \cdot X_i^{central}}{\sum_{i=1}^{j} f_i - 1}$$

Mediana (X_{MD}) – dados em ordem crescente

· Nº impar de observações – elemento central

· Nº par de observações – média dos elementos centrais

Para dados agrupados: regra de três

Moda (X_{MO}): elemento de maior ocorrência dentro do grupo

Amplitude (range): maior intervalo dentre os valores do conjunto de dados

$$A = [a, b]\,;\ a\ e\ b\ extremos$$

Variância amostral:

$$S_X^2 = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N - 1}$$

Para dados agrupados:

$$S_X^2 = \frac{\sum_{i=1}^{N} f_i \cdot (X_i^{central} - \bar{X})^2}{\sum_{i=1}^{j} f_i - 1}$$

__Inferência sobre conjunto de dados amostrais__ **(estimativas de parâmetros)**

Limites do intervalo de confiança

Se σ é conhecido:

$$\mu = \bar{X} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Se σ é desconhecido:

Distribuição t de Student com n-1 graus de liberdade

$$\mu = \bar{X} \pm t_{\alpha/2} \cdot \frac{S_X}{\sqrt{n}}$$

Proporção de sucessos $\hat{p}$ (amostral)

$$\hat{p} = \frac{nº\ de\ sucessos}{n} = \frac{X}{n}$$

X é o número de sucessos em n experim.

$$X = B(n, p) \cong N(np; np(1-p))$$

$$\hat{p} = N\left(p\,; \frac{p(1-p)}{n}\right)$$

Intervalo de confiança para p

$$p = \hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}}$$

Se não há nenhuma informação sobre p

$$p \approx \hat{p}$$

Se há informação do tipo $p \leq k$ (cte)

$$\begin{cases} p = 0{,}5 & se\ k \geq 0{,}5 \\ p = k & se\ k < 0{,}5 \end{cases}$$

Dados tabelados da normal $N(0, 1)$

$$\alpha = 90\% \implies z_{\alpha/2} = 1{,}65$$

$$\alpha = 95\% \implies z_{\alpha/2} = 1{,}96$$

$$\alpha = 99\% \implies z_{\alpha/2} = 2{,}58$$