Variáveis:

Qualitativos Nominal: a ordem das categorias não tem significado

EX: Género, Cor de olhos, Grupo Sanguíneo...

Qualitativos Ordinal: há uma ordem natural das categorias

Ex: Classe Social (Alta, baixa, Media), Nível da escola (1º, 2º, 3º ciclo) ...

Quantitativos Discreta: os valores podem ordenar-se, mas entre dois valores consecutivos não pode existir um valor intermedio (contagens)

Ex: nº cigarros, nº letras

Quantitativas Contínua: pode tomar qualquer valor num certo intervalo (medições)

Ex: Altura, peso

Tabela Frequências

Variáveis Qualitativas Nominais - não incluem as frequências acumuladas

Variáveis Qualitativas Ordinais ou Variáveis Quantitativas Discretas (com número pequeno de valores distintos) - incluem as frequências acumuladas

Variáveis Quantitativas Contínuas ou Variáveis Quantitativas Discretas (com número elevado de valores distintos) - Neste caso há a necessidade de agrupar os dados em classes e incluem as frequências acumulada

i	xi	ni	Ni	fi	Fi
Linha	Valor/Classe	Frequência absoluta (Contagem)	Frequência absoluta acumulada (∑ ni)	Freq relativa $(\frac{ni}{n})$	Freq Relativa Acumulado $\sum(\frac{ni}{n})$

Classes Para Tabela Frequências

Nº de classes (Regra de Sturges) $k = \lfloor 1 + log_2 n \rfloor = \left\lfloor 1 + \frac{\ln n}{\ln 2} \right\rfloor$

Amplitude dos dados $h = \frac{\max(xi) - \min(xi)}{k}$

Var Aleatória discreta

uma variável aleatória diz-se discreta se pode assumir um número finito ou infinito numerável de valores → associada a contagens

$f(x) = P(X = x)$ = Função de probabilidade

$f(x) \geq 0, \forall x$
$\sum f(x) = 1$
$F(x) = P(X \leq x)$ = Função de distribuição

$$F(x)\begin{cases} 0 \ , x \leq Min \\ \dots, Min \leq x < Max \\ 1, x \geq Max \end{cases}$$

$P(A < x \leq B) = f(B) - f(A)$ | $P(\overline{A}) = 1 - P(A)$ | $P(A|B) = \frac{P(A \cap B)}{P(B)}$

Símbolos $(\cup = OU\ ,\ \cap = E\ , | = Sabendo\ que)$

$E[X] = \mu = \Sigma\left(x * f(x)\right)$ = valor esperado

$E[X^2] = \Sigma\left(x^2 * f(x)\right)$

$E[aX + b] = a * E[X] + b$

$E[a] = a$

$E[X + Y] = E[X] + E[Y]$

$V[X] = E[X^2] - E^2[X]$ = variância = σ^2

$V[X] = E[(X - \mu)^2]$

$V[a] = 0$

$V[aX + b] = a^2 * V[X]$

$\sigma = \sqrt{V[X]}$ = Desvio Padrão

$E^2\,[X] = (E[X])^2$

Vars Aleatória discreta

Uniforme Discreta

Situações que todos os valores têm a mesma probabilidade de ocorrer

$X \sim U(n)$ $n = b - a + 1$ DX = {a, a + 1, a + 2, . . . , b}

$$f(x) = P(X = x) = \frac{1}{n} = \begin{cases} \frac{1}{n}, \forall\, x\; E\; Dx \\ 0, \qquad C.C \end{cases}$$

Caso o Dx seja inteiro consecutivos como Dx{1,2,3,4,5} (se possível tentar transformar em inteiro consecutivos)

$E[X] = \frac{a+b}{2}$ | $V[X] = \frac{(b-a+1)^2-1}{12}$

Se não

$E[X] = \frac{1}{n}\Sigma xi$ | $V[X] = \frac{1}{n}\Sigma xi^2 - (E[x])^2$

Binomial

Situações em que há 2 resultados possíveis (sucesso e insucesso)

$X \sim B(n, p)$ $n =$ provas de Bernoulli (número do domínio) , $p =$ probabilidade de sucesso

$E[X] = n * p$ | $V[X] = n * p * (1 - p)$

$f(x) = P(X = x) =$ dbinom (x, n, p)

$F(x) = P(X \leq x) =$ pbinom (x, n, p)

$F(k) = prob <=> P(X \leq k) = prob <=> k = F^{-1}(prob) =$ qbinom$(prob, n, p)$

Aditividade binomial $y = x_1 + x_2 \sim B(n, p)$ $n = nx_1 + nx_2, p = np_1 = np_2$

Poisson

Probabilidade de eventos num intervalo de tempo

$X \sim P[\lambda]$ | $E[xX] = V[X] = \lambda$

$f(x) = P(X = x) =$ dpois (x, λ)

$F(x) = P(X \leq x) =$ ppois (x, λ)

$F(k) = prob <=> P(X \leq k) = prob <=> k = F^{-1}(prob) =$ qpois$(prob, \lambda)$

Aditividade poisson $y = x_1 + x_2 \sim P(\lambda)$ $\lambda = \lambda x_1 + \lambda x_2$

Var Aleatória Continua

Uma variável aleatória diz-se contínua se pode assumir um número infinito não numerável de valores.

$F(x) = \int_{-\infty}^{x} f(t)dt \mid f(x) = \begin{cases} F'(x), Caso\ Exista \\ 0,\ C.C \end{cases}$

$f(x)$é uma e função densidade de probabilidade se satisfaz as seguintes propriedades

$f(x) \geq 0, \forall x\ \ \&\ \ \int_{-\infty}^{+\infty} f(x)dx = 1$

$F(x)$ pode ser $P(X < x)$ ou $P(X \leq x)$

$P(a \leq X \leq b) = \int_{a}^{b} f(x)\ dx \mid P(a \leq$ ou $< X \leq ou\ < b) = F(b) - F(a) \mid E[X] = \int_{-\infty}^{+\infty} x * f(x)\, dx$

$E[X^2] = \int_{-\infty}^{+\infty} x^2\, f(x)\, dx$

$E[a]\ = a$

$E[a * x + b]\ =\ a\ *\ E[x]\ + b$

$E^2\,[X] = (E[x])^2$

$V[X] =\ E[X^2] - E^2[X] =\ \sigma^2$ = Variância

$\sigma =\ \sqrt{\sigma^2} =\ \sqrt{V[X]}$ = Desvio Padrão

$V[a] = 0 \mid V[aX + b] = a^2 * V[X]$

Vars Aleatórias Continuas

Exponencial

Tempo ou distancia entre ocorrências sucessivas

$X{\sim}Exp(\theta)\ \ E[x] =\ \theta, V[x] =\ \theta^2$

Falta de memoria = $P(X \geq a + b\ | X \geq a) = P(X \geq b)$

$f(x) = P(X = x) =$ dunif (x, a, b)

$F(x) = P(X \leq x)\ ou\ P(X < x) =$ punif (x, a, b) = $\frac{x-a}{b-a}$

$F(k) = prob <=> P(X \leq k) = prob <=> k = F^{-1}(prob) =$ qunif$(prob, a, b)$

Uniforme Continua

Mesmo que a uniforme discreta, mas para valores contínuos

$X{\sim}U(a, b), Dx = [a, b], a =$1º elemento domínio $b =$ último elemento do domínio

$E[x] =\ \frac{a+b}{2}\ \ \mid\ V[x] = \frac{(b-a+1)^2-1}{12}$

$f(x) = P(X = x) =$ dexp $(x, \frac{1}{\theta})$

$F(x) = P(X \leq x)\ ou\ P(X < x) =$ pexp $(x, \frac{1}{\theta})$= $1 - e^{-\frac{x}{\theta}}, x \geq 0$

$F(k) = prob <=> P(X \leq k) = prob <=> k = F^{-1}(prob) =$ qexp$(prob, \frac{1}{\theta})$

NORMAL

$X \sim N(\mu, \sigma)$

$E[X] = \mu$, $V[X] = \sigma^2$

$f(x) = P(X = x) = \text{dnorm}\ (x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

$F(x) = P(X \leq x)\ ou\ P(X < x) = \text{pnorm}\ (x, \mu, \sigma)\ = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\, dt$

$F(k) = prob <=> P(X \leq k) = prob <=> k = F^{-1}(prob) = \text{qnorm}(prob, \mu, \sigma)$

Aditividade da normal

Sejam X1, X2, . . . , Xk variáveis aleatórias independentes com distribuição Normal, isto é

$Xi \sim N(\mu i, \sigma \mathrm{i}), \mathrm{i} = 1, \ldots..\mathrm{k}$ então

$$Y = \sum_{i=1}^{k} X_i \sim N\left(\sum_{i=1}^{k} \mu_i , \sqrt{\sum_{i=1}^{k} \sigma_i^2}\right)$$

Combinação Linear da Normal

Qualquer combinação linear de variáveis aleatórias independentes com distribuição Normal, ainda tem distribuição Normal, isto $Xi \sim N(\mu i, \sigma i), i = 1, \ldots..k$

$$Y = \sum_{i=1}^{k} a_i X_i \sim N\left(\sum_{i=1}^{k} a_i \mu_i , \sqrt{\sum_{i=1}^{k} a_i^2 \sigma_i^2}\right)$$

Normal Reduzida

diz-se que uma variável aleatória contínua Z tem distribuição Normal Reduzida (standard ou padrão) se a variável aleatória Z tem distribuição Normal com os parâmetros μ = 0 e σ = 1

$$\mathrm{Z} = \frac{\mathrm{X} - \mu}{\sigma} \sim \mathrm{N}(0{,}1)$$

$$\phi(z) = f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad z \in R$$

$$\Phi(\mathrm{z}) = \mathrm{F}(\mathrm{z}) = \mathrm{P}(\mathrm{Z} \leq \mathrm{z})\ \text{ou}\ \mathrm{P}(\mathrm{Z} < \mathrm{z}) = \int_{-\infty}^{\mathrm{z}} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{\mathrm{t}^2}{2}}\, \mathrm{dt}$$

$\Phi\ (-\mathrm{z})\ =\ 1\ -\ \Phi\ (\mathrm{z})$

$f(z) = P(Z = z) = \text{dnorm}(z)$

$F(z) = P(Z \leq z)\ ou\ P(Z < z) = \text{pnorm}\ (z)$

$F(k) = prob <=> P(Z \leq k) = prob <=> k = F^{-1}(prob) = \text{qnorm}(prob)$

Exemplo entre normal e normal reduzida

$x - Altura\ (\ldots) em\ m$

$\mu = ?$ $\quad X \sim N(\mu, 1.1) \quad P(X \geq 16) = 0.9$

$\sigma = 1.1$

Dados acima referidos dados no enunciado

$E[x] = ?$

$P(x \geq 16) = 0.9 <=> 1 - P(X < 16) = 0.9 <=> F(16) = 0.1 \quad (=) \left(Z < \frac{x-\mu}{\sigma}\right) \sim N(0,1)$

$P\left(Z < \frac{16-\mu}{\sigma}\right) = 0.1 <=> \Phi\left(\frac{16-\mu}{1.1}\right) = 0.1 <=> \frac{16-\mu}{1.1} = F^{-1}(0.1) =$ qnorm(0.1) = -1.282

$16 - \mu = 1.1 * -1.282 <=> \mu = 16 + 1.4102 = 17.41$

Amplitude Total

Habitualmente representa-se por A.

Para dados não agrupados, a amplitude total define-se como a diferença entre o maior e o menor valor do conjunto de dados (diferença entre os extremos). Isto e, seja $\{x1, x2, \ldots, xn\}$ um conjunto de dados com n observações,

$A = max(xi) - min(xi)$

Para dados agrupados em classes, a amplitude total é a diferença entre o limite superior da última classe e o limite inferior da primeira classe

Amplitude interquartis

A amplitude interquartil define-se como a diferença entre o 3 o quartil e o 1 o quartil:

$AIQ = Q3 - Q1 = Q0.75 - Q0.25$

Coeficiente de Variação

Para amostra

$CV = \frac{s}{\bar{x}} * 100\% \quad \bar{x} = media\ amostral \quad s = Desvio\ Padrão = \sqrt{(s^2)}$

$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - n\overline{x^2}}{n-1}$

Para População

$CV = \frac{\sigma}{\mu} * 100\%$

Caracterização da Distribuição de Frequências

(media = mediana = moda) é simétrica

(moda < mediana < média) é assimétrica positiva

(moda > mediana > média) é assimétrica Negativa

Medidas de Assimetria

$$b_1 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^3$$

$b1 = 0 =$ Simétrica | $b1 > 0$ =→ Assimétrica positiva | $b1 < 0 =$ Assimétrica negativa