

Anexo 1

Probabilidade e estatística aplicadas à hidrologia

Rita Cabral Guimarães

*ICAAM - Instituto de Ciências Agrárias e Ambientais Mediterrânicas,
Escola de Ciência e Tecnologia
Universidade de Évora*

1. Introdução

Nenhum processo hidrológico é puramente determinístico, isto é, não é possível determinar com exatidão a realização desse processo, pois ele está sujeito à ação de fatores aleatórios. Por exemplo, apesar de ser possível prever com alguma antecedência a ocorrência de precipitação, não é possível determinar qual a quantidade exata de precipitação que irá ocorrer.

Este facto parece estabelecer uma dificuldade básica no planeamento e gestão de qualquer sistema hidrológico, uma vez que para planear e gerir é fundamental conhecer o comportamento futuro dos processos que integram esse sistema hidrológico. No entanto, esta dificuldade pode ser ultrapassada, considerando que os processos hidrológicos são processos estocásticos, isto é, processos governados pelo menos em parte por fatores aleatórios. Se são processos estocásticos podem ser tratados recorrendo às leis de probabilidade e à estatística, sendo possível determinar qual a probabilidade duma realização desses processos se situar dentro de determinados intervalos. Por exemplo, se chover, pode determinar-se com antecedência qual probabilidade de ocorrer um determinado valor de precipitação.

2. Conceitos e definições

2.1. Frequência e probabilidade

Considere-se o lançamento de um dado perfeito. O conjunto de resultados possíveis desta experiência é conhecido e igual a $\Omega = \{1,2,3,4,5,6\}$. Chama-se experiência aleatória a uma experiência onde:

- É conhecido o conjunto Ω de todos os resultados possíveis;
- Não é possível conhecer, antes da realização da experiência, o resultado que ocorrerá (Lencastre e Franco, 2003).

Admita-se que se lança 20 vezes o dado e que a face 3 ocorre 5 vezes. A frequência (ou frequência relativa) de ocorrência da face 3, $f(3)$ é dada por:

$$f(3) = \frac{5}{20} = 0,25, \quad (\text{A1.1})$$

ou, genericamente,

$$f(x) = \frac{n}{N}, \quad (\text{A1.2})$$

onde n é o número de vezes em que ocorre o acontecimento x e N é o número de repetições da experiência (tamanho da amostra). Por exemplo (Hipólito e Vaz, 2011), se num registo de 50 anos de precipitação o acontecimento $x > 1200$ mm ocorrer 8 vezes na amostra, então a sua frequência relativa será,

$$f(> 1200) = \frac{8}{50} = 0,16.$$

2.2. População e amostra

Em estatística **população** designa um conjunto de elementos com alguma característica comum, por exemplo: os rios portugueses ou as precipitações anuais numa bacia hidrográfica. Pode dizer-se que a estatística se ocupa do estudo das propriedades das populações, populações estas que podem ser finitas ou infinitas conforme for finito ou infinito o número dos seus elementos. No entanto, e porque a observação de toda a população nem sempre é possível, o estudo das propriedades dessa população tem de ser feito sobre um seu subconjunto finito que se supõe ser representativo e se designa por **amostra**.

Quando, a partir da informação contida numa amostra, se tiram conclusões, expressas em termos de probabilidade, sobre toda a população entra-se no domínio da inferência estatística.

Considere-se uma amostra constituída por um determinado conjunto de dados x_1, x_2, \dots, x_n .

A diferença entre o maior e o menor dos valores dos dados, chama-se **amplitude dos dados**, I ,

$$I = \text{maior}X_i - \text{menor}X_i \quad (\text{A1.3})$$

Para resumir grandes quantidades de dados é usual distribuí-los em **classes**. O número de indivíduos pertencentes a cada classe denomina-se **frequência absoluta** da classe. A razão entre a frequência absoluta da classe e a frequência total (número total de valores da amostra) chama-se **frequência relativa** da classe.

À distribuição dos dados em classes com as respectivas frequências absolutas, chama-se **distribuição de frequências** ou **distribuição empírica** e à distribuição dos dados em classes com as respectivas frequências relativas, **chama-se distribuição de frequências relativas** ou **distribuição das percentagens** (Ver Exercício 1).

Geralmente, **o número de classes**, nc , deverá ser entre 5 e 20, no entanto, pode utilizar-se, para cálculo do número de classes, a fórmula sugerida por Sturges:

$$nc = 1 + \log_2 n = 1 + 3,32193 \log_{10} n . \quad (\text{A1.4})$$

Determinado o número de classes e uma vez conhecida a amplitude dos dados I , a amplitude de cada classe, c , pode ser determinada por:

$$c = \frac{I}{nc} \quad (\text{A1.5})$$

A representação gráfica duma distribuição de frequências fornece uma visão global da distribuição. Esta representação gráfica pode ser feita através de um **histograma**.

O histograma é uma sucessão de retângulos adjacentes, tendo cada um deles por base um segmento que corresponde à amplitude de cada classe e por altura as respectivas frequências absolutas ou relativas. (Ver Exercício 1)

3. Estatísticas descritivas de uma população e de uma amostra

Aspetos fundamentais para a caracterização das distribuições de frequência são as medidas de tendência central, medidas de dispersão e assimetria.

Às grandezas avaliadas a partir da população dá-se o nome de parâmetros e às grandezas calculadas com base na amostra dá-se o nome de estatísticas. Os parâmetros são representados por letras gregas (μ, σ, g) enquanto que as estatísticas são representadas por letras latinas (\bar{x}, S, g).

3.1. Medidas de tendência central

Considere-se uma amostra constituída por x_1, x_2, \dots, x_n onde n é o tamanho da amostra. Define-se:

3.1.1. Média ou valor médio

A média representa o centro de gravidade do sistema e é o mais importante parâmetro de localização. Designa-se por \bar{x} e define-se por,

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} . \quad (\text{A1.6})$$

3.1.2. Mediana

A mediana é o valor central da amostra ordenada por ordem crescente ($x_1 \leq x_2 \leq \dots \leq x_n$).

Assim, a mediana, M , pode definir-se por duas expressões:

- Se a amostra tem número ímpar de dados, $n = 2k + 1$, e a mediana vem,

$$M = x_{k+1}, \quad (\text{A1.7})$$

isto é a mediana é a observação central.

- Se a amostra tem número par de dados, $n = 2k$, e a mediana vem,

$$M = \frac{x_k + x_{k+1}}{2}, \quad (\text{A1.8})$$

isto é a mediana é a média dos dois valores centrais.

3.1.3. Moda

A moda é o valor mais frequente da amostra. É a medida de localização menos usada em hidrologia, pois em amostras de dados

hidrológicos (precipitações, caudais, etc.) é pouco provável que haja valores exatamente iguais. No entanto para cálculo da moda, Mod , pode utilizar-se a seguinte expressão,

$$Mod = \bar{x} - 3(\bar{x} - M). \quad (A1.9)$$

Onde \bar{x} e M são, respetivamente, a média e a mediana da amostra.

3.2. Medidas de dispersão

A dispersão pode definir-se como a posição dos dados em relação a uma referência fixa. Quando esta referência é a média, a dispersão indica o modo como os dados se espalham à volta do valor médio. Considere-se uma amostra constituída por x_1, x_2, \dots, x_n . Define-se:

3.2.1. Desvio padrão

O desvio padrão mostra o comportamento do conjunto de desvios em relação à média. Se a dispersão é grande, os desvios dos dados em relação à média são grandes e o desvio padrão será elevado. O contrário também se verifica quando os desvios são pequenos. O desvio padrão é dado por:

$$S = + \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}. \quad (A1.10)$$

Ao quadrado do desvio padrão, chama-se **variância**, S^2 ,

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}. \quad (A1.11)$$

3.2.2. Coeficiente de variação

O coeficiente de variação é um parâmetro adimensional que mede a variabilidade da amostra. Quanto maior o coeficiente de variação, maior é o desvio padrão em relação à média, isto é, mais dispersos estão os dados em torno da média. Define-se por,

$$C_v = \frac{S}{\bar{x}} \times 100\%. \quad (A1.12)$$

3.3. Assimetria

Assimetria é o grau de desvio, ou afastamento da simetria, de uma distribuição.

Quando se trabalha com distribuições de frequências, a assimetria pode ser estudada considerando a posição relativa dos três parâmetros de localização: média, mediana e moda.

Assim, nas distribuições **simétricas** (Figura A1.1), estes três parâmetros coincidem. Nas distribuições **assimétricas positivas** (desviadas para a direita) (Figura A1.2), **média>mediana>moda** e nas distribuições **assimétricas negativas** (desviadas para a esquerda) (Figura A1.3), **média<mediana<moda**.

A assimetria avalia-se pelo coeficiente de assimetria, g , sendo o valor deste coeficiente positivo nos desvios para a direita e negativo nos desvios para a esquerda.

$$g = \frac{n \sum_{i=1}^n (x_i - \bar{x})^3}{(n-1)(n-2)S^3}. \quad (\text{A1.13})$$

4. Distribuições de probabilidade

4.1. Variável aleatória.

Chama-se **variável aleatória** X a toda a variável suscetível de tomar diferentes valores de x^1 aos quais é possível afetar uma probabilidade. Chama-se **processo estocástico** a uma coleção ordenada de variáveis aleatórias $X_1, X_2, X_3, \dots, X_n$ e onde a sucessão cronológica $x_1, x_2, x_3, \dots, x_n$ resultante da sua observação, representa uma única realização do processo.

Uma variável aleatória diz-se **discreta** se só pode tomar um número finito de valores, por exemplo: o número de dias com chuva numa semana, mês ou ano, ou o número de vezes que o caudal ultrapassou determinado valor. Uma variável aleatória diz-se **contínua** se pode assumir qualquer valor dentro de um determinado intervalo de números reais, por exemplo: a precipitação anual, a temperatura média

¹ Para evitar confusões, a variável aleatória representa-se por maiúsculas, X , e as observações (ou realizações) dessa variável por minúsculas, x .

diária, etc., podem tomar qualquer valor dentro de um certo intervalo limitado por um mínimo e por um máximo.

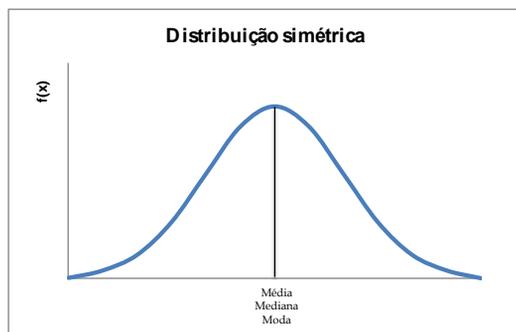


Figura A1.1. Distribuição simétrica (Média = Mediana = Moda)

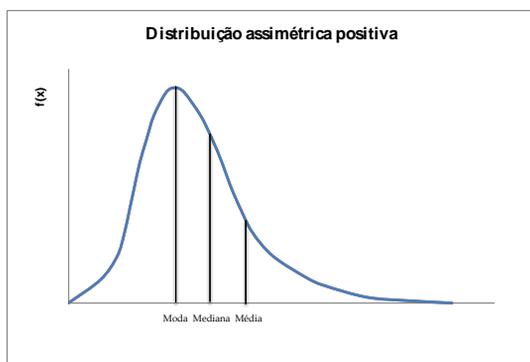


Figura A1.2. Distribuição assimétrica positiva (Média > Mediana > Moda).

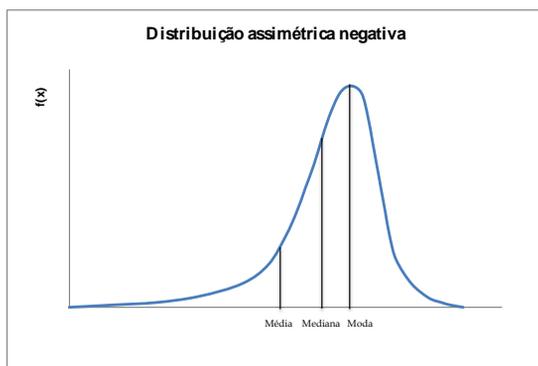


Figura A1.3. Distribuição assimétrica negativa (Média < Mediana < Moda).

4.2. Função de distribuição. Função densidade de probabilidade

Sendo X uma variável aleatória, dá-se o nome de **função de distribuição** (ou **função de distribuição de probabilidade**) da variável X à função,

$$F(x) = P(X \leq x), \quad (\text{A1.14})$$

que representa a probabilidade de a variável aleatória X assumir um valor inferior ou igual a x .

Ordenando por ordem **crecente** uma amostra de n valores duma variável aleatória, $x_1 \leq x_2 \leq \dots \leq x_n$, a probabilidade de a variável aleatória X assumir um valor inferior ou igual a x_i é:

$$F_a(x_i) = P(X \leq x_i) = \frac{m}{n}, \quad (\text{A1.15})$$

sendo $F_a(x_i)$ a **função de distribuição empírica** (FDE) da variável X e m o número de ordem do valor na amostra.

Ordenando por ordem **decrescente** a amostra, $x_1 \geq x_2 \geq \dots \geq x_n$, a probabilidade de a variável aleatória X assumir um valor superior ou igual a x_i é:

$$G_a(x_i) = P(X \geq x_i) = \frac{m}{n}, \quad (\text{A1.16})$$

sendo $G_a(x_i)$ a **função de duração da variável** X e m o número de ordem do valor na amostra.

Define-se **função densidade de probabilidade** $f(x)$ de uma variável aleatória contínua,

$$f(x) = \frac{dF(x)}{dx}, \quad (\text{A1.17})$$

$$f(x) = \text{Prob}\left(x - \frac{dx}{2} \leq X \leq x + \frac{dx}{2}\right). \quad (\text{A1.18})$$

4.3. Distribuições teóricas

Existem muitas **distribuições teóricas**, que servem como **modelo probabilístico** de variáveis ou fenómenos aleatórios. Considerando que as

variáveis hidrológicas são aleatórias, então elas podem ser representadas por algum tipo de distribuição teórica.

Apresentam-se de seguida as distribuições teóricas mais utilizadas em hidrologia.

4.3.1. Distribuições discretas

Distribuição binomial

É o modelo probabilístico indicado para descrever o número de sucessos em repetidas provas de Bernoulli.

As provas de Bernoulli (ou experiências de Bernoulli) são sucessões de experiências aleatórias independentes, onde em cada uma delas só existem dois resultados possíveis: realização de determinado acontecimento e realização do contrário desse acontecimento. Considerando um qualquer acontecimento, A , de probabilidade $P(A) = p$, a realização de, A , diz-se “sucesso” e a realização do contrário, \bar{A} , que tem probabilidade $P(\bar{A}) = 1 - p$, diz-se “insucesso”.

Por exemplo, a ocorrência de precipitação em determinado dia do futuro, só tem dois resultados possíveis: ou chove (sucesso) ou não chove (insucesso) nesse dia. Então, a probabilidade de chover é p , e a probabilidade de não chover, será logicamente $1-p$.

Se a **variável aleatória**, X , designar o **número de sucessos** em n provas, diz-se que tem **distribuição Binomial** e escreve-se simbolicamente $B(n, p)$. A sua função massa de probabilidade é,

$$P(x) = P(X = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n, \quad (\text{A1.19})$$

e a sua função de distribuição é,

$$F(x) = \sum_{x_i \leq x} \frac{n!}{x_i!(n-x_i)!} p^{x_i} (1-p)^{n-x_i}. \quad (\text{A1.20})$$

4.3.2. Distribuições contínuas

Distribuição normal

A mais importante e mais divulgada distribuição contínua de probabilidade é sem dúvida a **distribuição normal**. Teoricamente, a

função de distribuição da soma de n variáveis aleatórias tende para a distribuição normal quando n aumenta indefinidamente, qualquer que seja a função de distribuição de cada uma das variáveis aleatórias. Por esta razão a distribuição normal adapta-se bem a um grande número de variáveis hidrológicas, nomeadamente à **precipitação anual** e ao **escoamento anual**, resultantes da soma de um grande número de variáveis aleatórias.

Uma variável aleatória X com uma função densidade de probabilidade,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty < x < +\infty, \quad (\text{A1.21})$$

diz-se que tem **distribuição normal** com parâmetros μ e σ , e escreve-se simbolicamente, $N(\mu, \sigma)$. Os parâmetros μ e σ , são determinados por,

$$\mu = \bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad \text{e} \quad \sigma = S = +\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}.$$

A sua função de distribuição é dada por,

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx. \quad (\text{A1.22})$$

Para se efetuar o estudo da distribuição normal é necessário passar à **distribuição normal reduzida**, visto que os valores da função densidade de probabilidade e de distribuição são dados através de tabelas em função dos valores reduzidos. Isto consegue-se fazendo uma mudança de variável de modo a que a nova variável tenha valor médio igual a zero e desvio padrão igual à unidade. Isto é, transforma-se a variável X com $N(\mu, \sigma)$ numa variável Z com $N(0,1)$. Z é a variável reduzida, e é dada por,

$$Z = \frac{X - \mu}{\sigma}. \quad (\text{A1.23})$$

Ao realizar-se esta transformação, **standardiza-se a variável** X e neste caso a sua função densidade de probabilidade é,

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad -\infty < z < +\infty, \quad (\text{A1.24})$$

e a sua função de distribuição,

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{z^2}{2}} dz. \quad (\text{A1.25})$$

Os valores de $f(z)$ e $F(z)$ são dados por tabelas em função de z .

Na Figura A1.4 apresenta-se o gráfico da função densidade, $f(z)$, bem como os valores das ordenadas para os respectivos valores de z e na Figura A1.5 o gráfico da função distribuição, $F(z)$. A altura H na Figura A1.5 é a probabilidade acumulada correspondente à área tracejada na Figura A1.4.

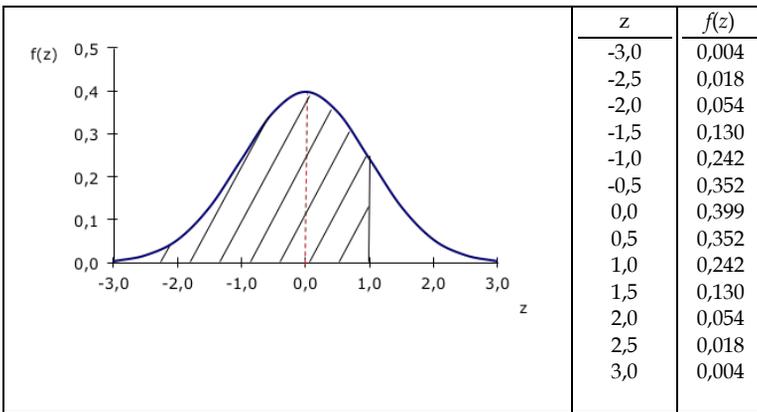


Figura A1.4. Função densidade de probabilidade $N(0,1)$

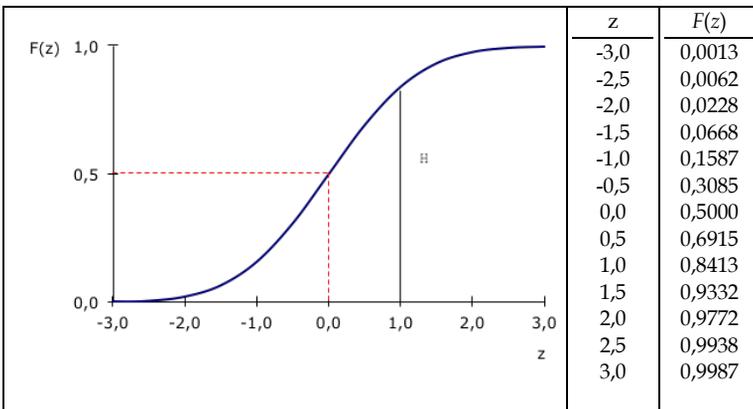


Figura A1.5. Função de distribuição de probabilidade $N(0,1)$

Como se pode verificar pelas Figuras A1.4 e A1.5, a distribuição normal é uma distribuição simétrica, isto é caracteriza-se por ter uma densidade de probabilidade simétrica em relação à média, que é ao mesmo tempo mediana e moda. Isto significa que a probabilidade média que a variável aleatória tem de se situar no intervalo $]-\infty, \mu]$ é igual à probabilidade média que ela tem de se situar no intervalo $[\mu, +\infty[$.

Como se pode observar na Figura A1.6 e comprovar com a tabela (Tabela 8.1), a área total limitada pela curva e pelo eixo dos x é unitária (100%).

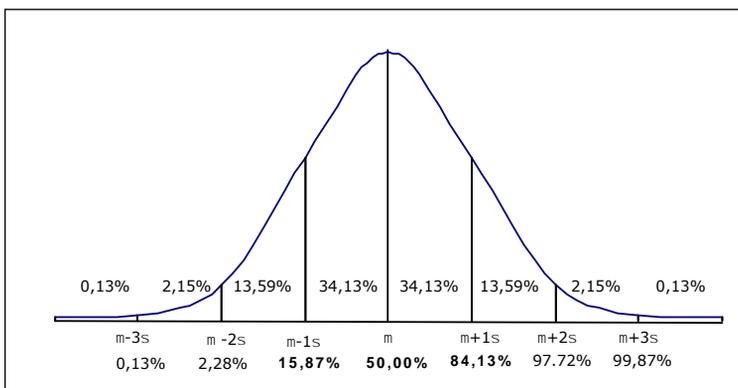


Figura A1.6. Áreas compreendidas pela curva normal reduzida (%).

Também se pode observar que 50% da distribuição normal corresponde ao intervalo $]-\infty, \mu]$. Isto significa que a probabilidade média de a variável aleatória se situar no intervalo $]-\infty, \mu]$ é igual 50% (0,13%+2,15%+13,59%+34,13%). Ou, por outras palavras, a probabilidade média de a variável aleatória ser igual ou inferior ao valor médio é 50%.

Também se pode observar na figura que 68,26% (34,13%+34,13%) da distribuição normal correspondem ao intervalo $[\mu \pm \sigma]$, que 95,44% (13,59%+34,13%+34,13%+13,59%) correspondem ao intervalo $[\mu \pm 2\sigma]$ e que 99,74% (2,25% + 13,59% + 34,13% + 34,13% + 13,59% + 2,15%) correspondem ao intervalo $[\mu \pm 3\sigma]$, significando que a probabilidade média da variável aleatória aumenta à medida que o intervalo alarga. (Ver Tabela 8.1).

Distribuição log-normal

Uma variável aleatória tem distribuição log-normal quando o logaritmo da variável aleatória tem distribuição normal. Isto é, se X segue a distribuição log-normal então $Y = \ln X$ segue a distribuição normal.

Se uma variável aleatória Y , tem distribuição normal, então a distribuição da variável X , diz-se log-normal, com função densidade de probabilidade e função de distribuição, respetivamente,

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu_y)^2}{2\sigma_y^2}} \quad x > 0, \quad (\text{A1.26})$$

onde, μ_y e σ_y são, respetivamente a média e o desvio padrão da variável

$$Y = \ln X, \text{ dados por } \mu_y = \bar{y} = \frac{\sum_{i=1}^n y_i}{n} \text{ e } \sigma_y = S_y = +\sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}}.$$

A distribuição log-normal ajusta-se bem a variáveis hidrológicas resultantes da multiplicação de muitas variáveis. Isto é, se $X = X_1 X_2 \dots X_n$, então $Y = \ln X = \sum_{i=1}^n \ln X_i = \sum_{i=1}^n Y_i$, que para n grande, tende para a distribuição normal.

Distribuição assintótica de extremos tipo I

Também conhecida por **distribuição de Gumbel**, é bastante aplicada a acontecimentos máximos, por exemplo, a distribuição dos caudais máximos anuais, ou a distribuição das precipitações máximas anuais.

Uma variável aleatória X , tem distribuição de Gumbel, com parâmetros α e u , quando a respetiva função densidade de probabilidade é da forma,

$$f(x) = \frac{1}{\alpha} e^{-\left[\frac{x-u}{\alpha} e^{\left(\frac{x-u}{\alpha} \right)} \right]} \quad -\infty < x < +\infty, \quad (\text{A1.27})$$

e a função distribuição é da forma,

$$F(x) = e^{-e^{\left(\frac{x-u}{\alpha} \right)}} \quad \alpha > 0. \quad (\text{A1.28})$$

Os parâmetros, α e u , podem ser determinados por,

$$\alpha = \frac{\sqrt{6}S_x}{\pi}, \quad (\text{A1.29})$$

e

$$u = \bar{x} - 0,5772\alpha. \quad (\text{A1.30})$$

Utilizando a variável reduzida de Gumbel,

$$y = \frac{x - u}{\alpha}, \quad (\text{A1.31})$$

vem para a função de distribuição,

$$F(x) = e^{-e^{-y}}. \quad (\text{A1.32})$$

No Quadro A1.1 apresentam-se, algumas das distribuições teóricas contínuas de probabilidade mais utilizadas em Hidrologia.

5. Distribuições teóricas e variáveis hidrológicas

5.1. Ajustamento de distribuições teóricas aos dados experimentais

Quando se afirma que as variáveis hidrológicas podem ser representadas por algum tipo conhecido de distribuição, não quer dizer que elas sigam perfeitamente essas distribuições teóricas. Obviamente que, quando se trata de variáveis reais, existem limitações, que tornam o ajuste perfeito impossível. Por exemplo, como já referido, a precipitação anual é uma variável que segue a distribuição normal. No entanto, a variável aleatória normal, pode assumir qualquer valor no intervalo $(-\infty, +\infty)$, enquanto que a precipitação apenas pode assumir valores positivos ou nulos. Além disso, como se viu, a distribuição normal é uma distribuição simétrica, enquanto que a distribuição de precipitação anual tende a ser assimétrica positiva.

Assim, quando se dispõe de uma amostra de valores de uma determinada variável hidrológica, o objetivo é determinar qual a **distribuição teórica que “melhor” se ajusta à distribuição empírica**. Depois de ajustar uma distribuição teórica conhecida a um conjunto de variáveis hidrológicas, grande parte da informação probabilística da

amostra pode ser descrita por essa distribuição teórica e pelos respectivos parâmetros.

O ajustamento de uma distribuição teórica à distribuição empírica de variáveis hidrológicas, pode ser efetuado através de **testes de hipóteses** estatísticos ou através do **posicionamento gráfico**.

Quadro A1.1 - Distribuições teóricas contínuas utilizadas em Hidrologia

Distribuição	F. densidade de Probabilidade	Intervalo	Equações dos parâmetros
Normal	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$-\infty < x < +\infty$	$\mu = \bar{x}, \sigma = S_x$
Log - normal	$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu_y)^2}{2\sigma_y^2}}$ $y = \ln x$	$x > 0$	$\mu_y = \bar{y}, \sigma_y = S_y$
Pearson Tipo III	$f(x) = \frac{\lambda^\beta (x-\varepsilon)^{\beta-1} e^{-\lambda(x-\varepsilon)}}{\Gamma(\beta)}$	$x \geq \varepsilon$	$\beta = \left(\frac{2}{g}\right)^2, \lambda = \frac{S_x}{\sqrt{\beta}}$ $\varepsilon = \bar{x} - S_x\sqrt{\beta}$
Log - Pearson	$f(x) = \frac{\lambda^\beta (y-\varepsilon)^{\beta-1} e^{-\lambda(y-\varepsilon)}}{\Gamma(\beta)}$ $y = \ln x$	$\ln x \geq \varepsilon$	$\beta = \left(\frac{2}{g_y}\right)^2, \lambda = \frac{S_y}{\sqrt{\beta}}$ $\varepsilon = \bar{y} - S_y\sqrt{\beta}$
Gumbel	$f(x) = \frac{1}{\alpha} e^{\left[-\frac{x-u}{\alpha} - e^{-\frac{x-u}{\alpha}} \right]}$	$-\infty < x < +\infty$	$\alpha = \frac{\sqrt{6}S_x}{\pi},$ $u = \bar{x} - 0,5772\alpha$

5.1.1. Testes de hipóteses

Quando se pretende saber se uma determinada variável aleatória segue uma qualquer distribuição teórica, utiliza-se um teste de hipóteses.

O estabelecimento de um teste de hipóteses segue os seguintes passos:

- 1º - Formulação da hipótese a ser testada, H_0 - Hipótese nula;
- 2º - Formulação da hipótese alternativa, H_1 ;
- 3º - Seleção da estatística amostral a ser utilizada;

4º - Estabelecimento da regra de decisão, em função de uma constante c ;

5º - Seleção do nível de significância, α ;

6º - Utilização da estatística amostral para determinar o valor da constante c , de modo a que, quando H_0 for verdadeira, haja uma probabilidade α de se rejeitar esta hipótese;

7º - Rejeição ou não rejeição da hipótese H_0 , se a estatística amostral observada cair, respetivamente, na região de rejeição (crítica), ou na região de não rejeição.

Teste do qui-quadrado

O teste do qui-quadrado, χ^2 , é um teste de adequação do ajustamento, onde se pretende determinar se uma dada distribuição teórica é razoável face aos dados disponíveis. Assim, as hipóteses a testar são,

H_0 : A função de distribuição é $F(x)$;

H_1 : A função de distribuição **não** é $F(x)$.

O teste do qui-quadrado, faz uma comparação entre o número real de observações e o número esperado de observações que caem nas respetivas classes, através do cálculo da estatística,

$$\chi_c^2 = \sum_{j=1}^{nc} \frac{(O_j - E_j)^2}{E_j}, \quad (\text{A1.33})$$

que assintoticamente tem distribuição de qui-quadrado com $\nu = m - p - 1$ graus de liberdade, sendo nc o número de classes, p o número de parâmetros a estimar a partir da amostra, O_j o número de observações na classe j , e E_j o número de observações que seriam de esperar, na classe j , através da distribuição teórica.

A decomposição da amostra em classes, deve ser tal que o efetivo teórico por classe não seja inferior a 5, ou pode ser utilizada a fórmula de Sturges (Equação A1.4).

As classes devem ser escolhidas de modo a que a cada intervalo de classe corresponda uma probabilidade igual, (classes equiprováveis), donde,

$$E_j = \frac{n}{nc}. \quad (\text{A1.34})$$

A hipótese H_0 é rejeitada se χ_c^2 for maior que $\chi_{1-\alpha, \nu}^2$ tabelado, para um determinado nível de significância α e ν graus de liberdade. (Tabela 8.2).

Teste de Kolmogorov-Smirnov

Uma alternativa ao teste do χ^2 , é o teste de Kolmogorov-Smirnov. Para a realização deste teste, deve considerar-se:

1º $F(x)$ a função teórica da distribuição acumulada admitida como hipótese nula, H_0 ;

2º $F_0(x)$ a função de distribuição acumulada para os dados amostrais $\left(\frac{m}{n}\right)$;

3º A estatística utilizada,

$$D = \max|F(x) - F_0(x)|; \quad (\text{A1.35})$$

4º Se, para um determinado nível de significância α , o valor D calculado pela Equação (A1.35) for maior ou igual ao valor D tabelado a hipótese H_0 é rejeitada.

5.1.2. Posicionamento gráfico

Para avaliar o ajustamento de uma distribuição teórica à distribuição empírica dos dados amostrais pode, também, recorrer-se ao **posicionamento gráfico dos dados** na forma de uma distribuição cumulativa de probabilidade.

Considere-se uma amostra $x_1, x_2, x_3, \dots, x_n$, atribuindo a estes dados amostrais, uma probabilidade empírica $F(x) = P(X \leq x)$ ou $G(x) = P(X > x)$, é possível marcar estes pares de valores $[x, F(x)]$ ou $[x, G(x)]$ num gráfico de eixos coordenados.

A função de distribuição de uma determinada distribuição teórica pode ser representada graficamente num **papel de probabilidade** adequado a essa distribuição. Em tal papel, as ordenadas representam os valores da variável X e as abcissas representam a probabilidade de não excedência $F(x) = P(X \leq x)$, a probabilidade de excedência

$G(x) = P(X > x)$, o período de retorno T , ou a variável reduzida Y . As escalas das ordenadas e das abcissas são feitas de tal modo, que a função de distribuição teórica aparece representada por uma reta. Sendo assim, se os dados amostrais, afetados da respetiva probabilidade empírica, **se ajustam à reta da distribuição teórica**, então pode afirmar-se que a **distribuição empírica segue a distribuição teórica considerada**.

Para afetar cada valor dos dados amostrais da respetiva probabilidade, suponha-se que se dispõe de todas as observações de uma variável aleatória. Se as n observações x forem classificadas por ordem crescente, a **probabilidade empírica** de X tomar valores inferiores ou iguais a um determinado x_i será:

$$F(x) = (P(X \leq x_i)) = \frac{m}{n}, \quad (\text{A1.36})$$

onde m é o número de ordem do valor na amostra. Se as n observações x forem classificadas por ordem decrescente, a probabilidade empírica de X tomar valores iguais ou superiores a um determinado x_i será:

$$G(x) = (P(X > x_i)) = \frac{m}{n}. \quad (\text{A1.37})$$

Utilizando estas fórmulas, o menor valor da população teria uma probabilidade igual a zero e o maior valor uma probabilidade igual a um. No entanto, a afetação de probabilidade a uma amostra é mais delicada, pois não há a certeza de que ela contenha o menor e o maior valor da população desconhecida. Das várias fórmulas existentes para afetar cada valor da amostra de uma probabilidade empírica, utilizar-se-á a de Weibull, por ser a mais generalizada,

$$F_a(x) = (P(X \leq x_i)) = \frac{m}{n+1}, \quad (\text{A1.38})$$

para os n dados classificados **por ordem crescente** e

$$G_a(x) = (P(X > x_i)) = \frac{m}{n+1}, \quad (\text{A1.39})$$

para os n dados classificados **por ordem decrescente**.

6. Análise frequencial em hidrologia

Nos sistemas hidrológicos existem muitas vezes eventos extremos, tais como secas ou cheias. O valor de um acontecimento extremo é inversamente proporcional à sua frequência de ocorrência, isto é, um acontecimento extremo ocorre com menor frequência do que um evento moderado. O objetivo da **análise frequencial** em hidrologia é relacionar a **magnitude dos valores extremos com a sua frequência de ocorrência**, através da utilização de distribuições de probabilidade. Os resultados desta análise podem ser usados em vários problemas de engenharia, tais como, dimensionamento de barragens, pontes, estruturas de controlo de cheias, etc..

Para efetuar a análise frequencial pode recorrer-se ao **posicionamento gráfico** dos dados na forma de uma distribuição cumulativa de probabilidade ou utilizar técnicas analíticas baseadas em **fatores de frequência**.

Em qualquer dos casos torna-se necessário introduzir a noção de período de retorno.

6.1. Período de retorno. Risco hidrológico

O **período de retorno** de uma variável X , define-se como o número de anos que deve, em média, decorrer para que o valor dessa variável ocorra ou seja superado.

O **período de retorno**, T , pode exprimir-se por,

$$T = \frac{1}{G(x)} = \frac{1}{1-F(x)}, \quad (\text{A1.40})$$

onde $F(x)$ é a **probabilidade de não excedência** dada por $F(x) = P(X \leq x)$ e $G(x)$ é a **probabilidade de excedência** dada por $G(x) = P(X > x) = 1 - P(X \leq x) = 1 - F(x)$.

Risco hidrológico, R , é função do período de retorno e representa a probabilidade de um valor x da variável aleatória X ser excedido em pelo menos uma vez em n anos sucessivos. Exprime-se por,

$$R = \frac{n}{T} = 1 - \left(1 - \frac{1}{T}\right)^n = 1 - (1 - G(x))^n \quad (\text{A1.41})$$

6.2. Análise frequencial por posicionamento gráfico

Depois de verificado o ajustamento de uma distribuição teórica aos dados amostrais podemos considerar que a distribuição amostral é representada pela distribuição teórica. Assim, para efetuar a análise frequencial, basta retirar da reta teórica os valores da variável aleatória e as respetivas probabilidades, como veremos no Exercício 7.

6.3. Análise frequencial por fatores de frequência

A análise frequencial pode ser feita recorrendo a técnicas analíticas baseadas em **fatores de frequência**.

Chow *et al.* (1988) propõem a seguinte fórmula geral para a análise hidrológica de frequências,

$$x_T = \bar{x} + K_T S, \quad (\text{A1.42})$$

onde, x_T é o valor do acontecimento associado a determinado período de retorno, \bar{x} é a média da amostra, S é o desvio padrão e K_T é o **fator de frequência** que é função do período de retorno, T , e do tipo de distribuição de probabilidade a ser utilizada na análise.

Se a variável em análise é $y = \ln x$, o mesmo método pode ser utilizado, aplicado aos logaritmos dos dados,

$$y_T = \bar{y} + K_T S_y. \quad (\text{A1.43})$$

O fator de frequência proposto por Ven Te Chow é aplicável a muitas distribuições de probabilidade utilizadas na análise hidrológica de frequências. Para uma determinada distribuição teórica, é possível determinar uma relação entre o fator de frequência e o correspondente período de retorno ($K-T$), relação esta que pode ser expressa por tabelas ou analiticamente.

Para determinar o valor de x_T (Equação A1.42), é necessário calcular os parâmetros estatísticos para a distribuição proposta e determinar para um dado período de retorno, o fator de frequência.

Seguidamente descreve-se a relação teórica $K-T$, para várias distribuições de probabilidade.

a) Distribuição normal - O fator de frequência pode ser expresso por,

$$K_T = \frac{x_T - \mu}{\sigma} = z, \quad (\text{A1.44})$$

que é a mesma expressão da variável normal reduzida Z , definida na equação (A1.24).

b) Distribuição log-normal - Para a distribuição log-normal o fator de frequência pode ser expresso por,

$$K_T = \frac{y_T - \mu_y}{\sigma_y}, \quad (\text{A1.45})$$

onde $y = \ln x$. Este fator de frequência aplica-se à equação (A1.43).

c) Distribuição de Gumbel - Para esta distribuição, o fator de frequência é determinado por,

$$K_T = -\frac{\sqrt{6}}{\pi} \left\{ 0,5772 + \ln \left[\ln \left(\frac{T}{T-1} \right) \right] \right\}. \quad (\text{A1.46})$$

Para expressar T , em termos de K_T , utiliza-se a seguinte equação,

$$T = -\frac{1}{1 - e^{-e^{-\left[0,5772 - \frac{\pi K_T}{\sqrt{6}} \right]}}}. \quad (\text{A1.47})$$

7. Exercícios resolvidos

Exercício 1

Considerem-se as precipitações anuais registadas na estação de Castro D'Aire durante 39 anos, apresentadas no Quadro A1.2. Calcule:

- A distribuição de frequências, a distribuição de frequências relativas da precipitação anual e o histograma de frequências absolutas;
- A média, a mediana e a moda;
- O desvio padrão e o coeficiente de variação;
- O coeficiente de assimetria.

Quadro A1.2. Precipitação anual (mm) em Castro D' Aire

Ano Hidrológico	Precipitação (xi) [mm]	Precipitação ordenada de forma crescente [mm]	Nº de ordem (i)
1916/17	2118,2	912,8	1
1917/18	1001,2	915,8	2
1918/19	2093,2	1001,2	3
1919/20	1556,4	1076,1	4
1920/21	1290,6	1127,1	5
1921/22	1785,4	1201,0	6
1922/23	1830,2	1254,0	7
1923/24	2150,2	1275,7	8
1924/25	1749,6	1290,6	9
1925/26	2221,6	1301,2	10
1926/27	2024,1	1334,4	11
1927/28	1923,7	1344,7	12
1928/29	1127,1	1411,7	13
1929/30	2630,9	1461,0	14
1930/31	1481,2	1481,2	15
1931/32	1461,0	1520,6	16
1932/33	1334,4	1556,4	17
1933/34	1301,2	1581,0	18
1934/35	1581,0	1625,0	19
1935/36	3249,6	1664,2	20
1936/37	2069,0	1699,5	21
1937/38	1254,0	1749,6	22
1938/39	1974,0	1763,0	23
1939/40	2059,6	1785,4	24
1940/41	2569,6	1830,2	25
1941/42	1520,6	1903,9	26
1942/43	1664,2	1923,7	27
1943/44	1344,7	1974,0	28
1944/45	915,8	2024,1	29
1945/46	1763,0	2059,6	30
1946/47	2079,3	2069,0	31
1947/48	1411,7	2079,3	32
1948/49	912,8	2093,2	33
1949/50	1201,0	2118,2	34
1950/51	1903,9	2150,2	35
1951/52	1625,0	2221,6	36
1952/53	1076,1	2569,6	37

1953/54	1275,7	2630,9	38
1954/55	1699,5	3249,6	39

Resolução:

a) A amplitude dos dados determina-se facilmente pela equação (A1.3):

$$I = 3249,6 - 912,8 = 2336,8 \text{ mm,}$$

o número de classes, utilizando a equação (A1.4), é:

$$nc = 1 + 3,32193 \log_{10} 39 = 6 \text{ classes,}$$

e a amplitude de cada classe, determina-se recorrendo à equação (A1.5):

$$c = \frac{2336,8}{6} = 389,5 \text{ mm}$$

Isto é, a 1ª classe terá como limite inferior o valor 912,8 mm e como limite superior 1302,3 mm (912,8 + 389,5), a 2ª classe terá como limite inferior 1302,3 mm e como limite superior 1691,5 mm (1302,3 + 389,5), e assim sucessivamente até ao limite superior da última classe.

Para determinar os valores pertencentes a cada classe, que conduz às frequências absolutas e relativas de cada classe, basta contar os elementos que caem em cada classe, considerando que um **determinado valor x pertence a uma classe** quando e só quando é **maior que o limite inferior e menor ou igual que o limite superior dessa classe**.

A divisão da amostra em classes bem como as frequências absolutas e relativas de cada classe são apresentadas no Quadro A1.3.

Quadro A1.3. Distribuição de frequências e distribuição de frequências relativas da precipitação anual em Castro D'Aire.

Precipitação anual (mm)	Frequências absolutas	Frequências relativas
912,8 - 1302,3	10	0,256410256
1302,3 - 1691,8	10	0,256410256
1691,8 - 2081,3	12	0,307692308
2081,3 - 2470,8	4	0,102564103
2470,8 - 2860,3	2	0,051282051
2860,3 - 3249,8	1	0,025641026
TOTAL	39	1

Na Figura A1.7 apresenta-se o histograma das frequências absolutas.

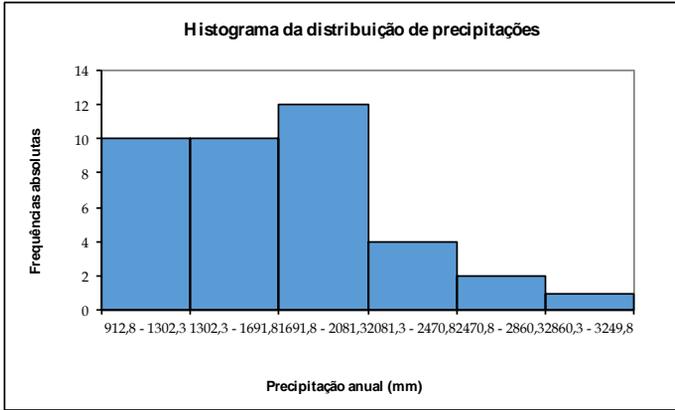


Figura A1.7. Histograma das frequências absolutas para a precipitação anual em Castro D'Aire.

b) Utilizando as equações (A1.6), (A1.7) e (A1.9) vem, respetivamente, para a média, mediana e moda:

$$\bar{x} = \frac{2118,2 + \dots + 1699,5}{39} = \frac{66230,3}{39} = 1698,2 \text{ mm}$$

$$n = 39 \Rightarrow k = \frac{38}{2} = 19$$

$$M = x_{k+1} = x_{20} = 1664,2 \text{ mm}$$

$$Mod = 1698,2 - 3(1698,2 - 1664,2) = 1596,2 \text{ mm}$$

Para esta distribuição de precipitações, tem-se que, $\bar{x} > M > Mod$.

c) Utilizando as equações (A1.10) e (A1.12) vem, respetivamente, para o desvio padrão e coeficiente de variação:

$$S = \sqrt{\frac{(2118,2 - 1698,2)^2 + \dots + (1699,5 - 1698,2)^2}{39 - 1}} = 498,9 \text{ mm};$$

$$C_v = \frac{498,9}{1698,2} \times 100 = 29,4\%$$

d) Utilizando a equação (A1.13) vem para o coeficiente de assimetria:

$$g = \frac{39[(2118,2 - 1698,2)^3 + \dots + (1699,5 - 1698,2)^3]}{(39-1)(39-2) \times 498,9^3} = 0,8.$$

Como a distribuição tem assimetria positiva, significa que $\bar{x} > M > Mod$ (estatísticas já determinadas na alínea b), isto é, trata-se de uma distribuição desviada para a direita.

Exercício 2

Considerando que em determinado rio ocorre uma cheia por ano e que a probabilidade desta cheia ser catastrófica é 10%, qual é a probabilidade de ocorrência de 3 destas cheias nos próximos 15 anos?

Resolução:

Neste caso, tem-se,

$$n = 15 \text{ anos,}$$

$$x = 3,$$

$$p = 0,1,$$

logo, pela equação (A1.19) vem,

$$P(3) = P(X = 3) = \frac{15!}{3!(15-3)!} 0,1^3 (1-0,1)^{15-3} = 0,1285$$

Isto é, nos próximos 15 anos a probabilidade de ocorrência de 3 cheias catastróficas neste rio é de 12,85%.

Exercício 3

Admitindo que a precipitação anual em determinado local, é uma variável aleatória X , com distribuição normal e com parâmetros $\mu = 570mm$ e $\sigma = 120mm$, $N(570;120)$, determinar a probabilidade de um valor de precipitação ser menor que 600 mm.

Resolução:

Queremos determinar $F(x)$ que corresponde a $x \leq 600mm$. Para tal, devemos transformar a variável X com $N(570;120)$ na variável reduzida Z com $N(0,1)$, isto é, devemos, utilizando a Equação A1.23, calcular:

$$z = \frac{600 - 570}{120} = 0,25.$$

Assim, já podemos recorrer à Tabela 8.1 e retirar o valor da probabilidade pretendida. Pela tabela vem para $z = 0,25$ uma probabilidade $F(z) = 0,5987$. Isto é probabilidade de a variável X , assumir um valor $x \leq 600\text{mm}$ é de 59,87%.

Exercício 4

Considerando que o caudal anual de determinado curso de água, é uma variável aleatória X , com distribuição log-normal, com $\mu_y = 5,06146$ e $\sigma_y = 0,58906$, determinar a probabilidade de se verificar um valor de caudal inferior a $x \leq 150 \text{ m}^3 \text{ s}^{-1}$.

Resolução:

Fazendo uma mudança na variável, tal que $Y = \ln X$, vem, $y = \ln x = \ln 150 = 5,01064$, donde a variável reduzida Z (Equação A1.23) é,

$$z = \frac{5,01064 - 5,06146}{0,58906} = -0,09.$$

Para obter o correspondente valor de $F(z)$, utiliza-se a Tabela 8.1. Para $z = -0,09$ vem, $F(z) = 1 - F(z) = 1 - 0,5359 = 0,4641$. Que significa que a probabilidade de se verificar um valor de caudal inferior a $x \leq 150 \text{ m}^3 \text{ s}^{-1}$ é de 46,41%.

Exercício 5

Os caudais máximos instantâneos anuais num determinado curso de água seguem a distribuição de Gumbel, com média, $\bar{x} = 227 \text{ m}^3 \text{ s}^{-1}$ e desvio padrão, $S = 142 \text{ m}^3 \text{ s}^{-1}$. Determine a probabilidade de ocorrer um valor de caudal $x \leq 300 \text{ m}^3 \text{ s}^{-1}$.

Resolução:

Os parâmetros, α e μ , podem ser determinados pelas equações (A1.29) e (A1.30),

$$\alpha = \frac{\sqrt{6}S_x}{\pi} = \frac{\sqrt{6} \times 142}{\pi} = 110,7 \text{ m}^3 \text{ s}^{-1},$$

$$u = \bar{x} - 0,5772\alpha = 227 - 0,5772 \times 110,7 = 163,1 \text{ m}^3 \text{ s}^{-1}.$$

Utilizando a variável reduzida de Gumbel (Equação A1.31),

$$y = \frac{300 - 163,1}{110,7} = 1,237,$$

podemos determinar a probabilidade pretendida por aplicação da equação (A1.32),

$$F(x) = e^{-e^{-1,237}} = 0,748 = 74,8\%.$$

Exercício 6

Verificar o ajustamento das precipitações anuais ocorridas na estação meteorológica de Castro D'Aire (Quadro A1.2) à distribuição normal.

Resolução:

Esta verificação pode ser feita de duas maneiras: por **posicionamento gráfico dos dados** ou através de **um teste de adequação do ajustamento**.

a) Posicionamento gráfico - Num papel de probabilidade normal, qualquer distribuição normal terá como gráfico uma linha reta, correspondendo a média dessa distribuição ao ponto 50% e um desvio padrão para cada lado da média, aos pontos 15.87% e 84.13%, respetivamente (ver Figura A1.6).

Neste caso, a média e o desvio padrão foram determinados no Exercício 1 alínea a) e b) e são, respetivamente, 1698,2 mm e 498,9 mm. A reta da distribuição normal teórica desenha-se no papel normal unindo os três pares de pontos,

$$(\bar{x} - S; 15,87\%) = (1199,3; 15,87\%)$$

$$(\bar{x}; 50\%) = (1698,2; 50\%)$$

$$(\bar{x} + S; 84,13\%) = (2197,1; 84,13\%)$$

Esta reta corresponde à distribuição normal teórica. Se os valores da amostra, afetados da respetiva probabilidade empírica, ajustarem à

reta, então pode afirmar-se que a série de precipitações anuais segue a distribuição normal.

Para atribuir uma probabilidade empírica aos valores da amostra, utiliza-se a fórmula de Weibull, dada pela Equação (A1.38),

$$F_a(x) = \frac{m}{n+1}$$

que dá a probabilidade de não excedência, $F(x)$, para os n valores da amostra, ordenados de forma crescente. Quadro A1.4.

Quadro A1.4. Probabilidade de não excedência, $F(x)$, para os valores de precipitação anual em Castro D'Aire.

Ano	Precipitação (x_i) [mm]	Precipitação ordenada de forma crescente [mm]	Nº de ordem (i)	$F(x)$ [%]
1916/17	2118,2	912,8	1	2,5
1917/18	1001,2	915,8	2	5,0
1918/19	2093,2	1001,2	3	7,5
1919/20	1556,4	1076,1	4	10,0
1920/21	1290,6	1127,1	5	12,5
1921/22	1785,4	1201,0	6	15,0
1922/23	1830,2	1254,0	7	17,5
1923/24	2150,2	1275,7	8	20,0
1924/25	1749,6	1290,6	9	22,5
1925/26	2221,6	1301,2	10	25,0
1926/27	2024,1	1334,4	11	27,5
1927/28	1923,7	1344,7	12	30,0
1928/29	1127,1	1411,7	13	32,5
1929/30	2630,9	1461,0	14	35,0
1930/31	1481,2	1481,2	15	37,5
1931/32	1461,0	1520,6	16	40,0
1932/33	1334,4	1556,4	17	42,5
1933/34	1301,2	1581,0	18	45,0
1934/35	1581,0	1625,0	19	47,5
1935/36	3249,6	1664,2	20	50,0
1936/37	2069,0	1699,5	21	52,5
1937/38	1254,0	1749,6	22	55,0
1938/39	1974,0	1763,0	23	57,5
1939/40	2059,6	1785,4	24	60,0
1940/41	2569,6	1830,2	25	62,5
1941/42	1520,6	1903,9	26	65,0
1942/43	1664,2	1923,7	27	67,5
1943/44	1344,7	1974,0	28	70,0

1944/45	915,8	2024,1	29	72,5
1945/46	1763,0	2059,6	30	75,0
1946/47	2079,3	2069,0	31	77,5
1947/48	1411,7	2079,3	32	80,0
1948/49	912,8	2093,2	33	82,5
1949/50	1201,0	2118,2	34	85,0
1950/51	1903,9	2150,2	35	87,5
1951/52	1625,0	2221,6	36	90,0
1952/53	1076,1	2569,6	37	92,5
1953/54	1275,7	2630,9	38	95,0
1954/55	1699,5	3249,6	39	97,5

A reta teórica de probabilidade normal e os valores da distribuição empírica da precipitação anual estão representados na Figura A1.8, onde se pode verificar, qualitativamente, o ajustamento à reta, donde se pode afirmar que **a série de precipitações em estudo tem distribuição normal**.

b) Teste do qui-quadrado - Para melhor ajuizar da qualidade do ajustamento da distribuição normal à distribuição empírica de precipitações anuais, utiliza-se o teste de hipótese do χ^2 .

As hipóteses a testar são,

H_0 : A função de distribuição é normal;

H_1 : A função de distribuição não é normal.

O número de classes, nc , para esta amostra é 6 (determinado no Exercício 1 alínea a).

Uma vez que é necessário trabalhar com as tabelas para a distribuição normal, utilizar-se-á a variável reduzida z .

Como as classes devem ser equiprováveis vem para a probabilidade de cada classe,

$$F(z) = \frac{1}{6} = 0,1667.$$

Os z_i serão calculados, a partir dos valores $F(z_i)$ conhecidos, por consulta da Tabela 8.1. A partir de z_i determinam-se facilmente os intervalos das classes, x_i , sabendo que $\bar{x} = 1698,2$ mm e $S = 498,9$ mm. Como se mostra no Quadro A1.5.

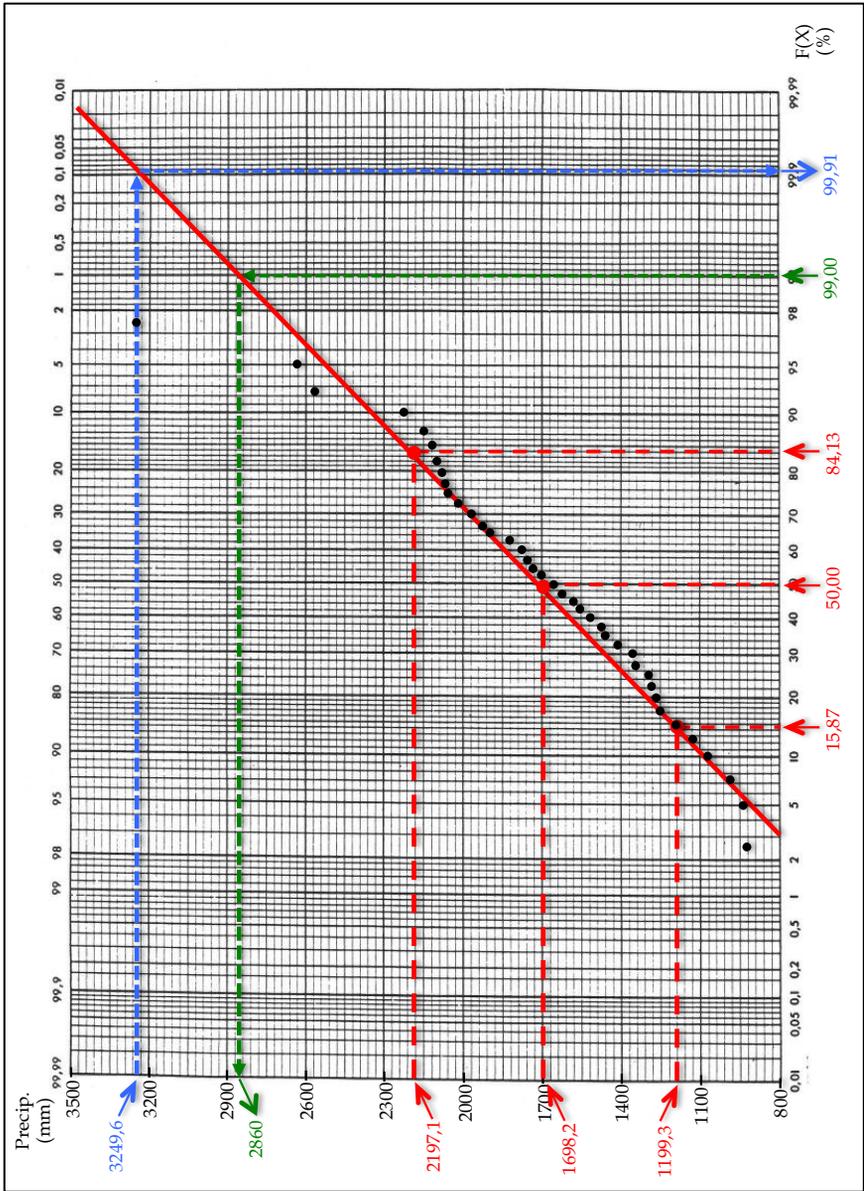


Figura A1.8. Ajustamento da distribuição normal aos valores de precipitação anual em Castro D'Aire.

Quadro A1.5. Cálculo dos intervalos e limites das classes

z_i	$F(z_i)$	z_i	$x_i = z_i \times S + \bar{x}$
z_1	$1/6 = 0,1667$	-0,9674	1215,6
z_2	$2/6 = 0,3333$	-0,4307	1483,3
z_3	$3/6 = 0,5000$	0,0000	1698,2
z_4	$4/6 = 0,6667$	0,4307	1913,1
z_5	$5/6 = 0,8333$	0,9674	2180,8

Com estes elementos pode construir-se o Quadro A1.6 e calcular o χ_c^2 pela Equação (A1.33).

Da Tabela 8.2 vem, para $\alpha = 0,05$ e $\nu = 6 - 2 - 1 = 3$ graus de liberdade,

$$\chi_{0,95;4}^2 = 7,81 .$$

Como $\chi_c^2 = 2,6923 < \chi_{0,95;4}^2 = 7,81$ pode dizer-se que a **hipótese de normalidade não é rejeitada**, o que vem confirmar a análise gráfica feita na alínea anterior.

Quadro A1.6 Teste do qui-quadrado

Limite das classes	Nº de elementos esperados em cada classe (E_i)	Nº de elementos observados em cada classe (O_i)	
$\leq 1215,6$	6,5000	7	0,0385
1215,6 - 1483,3	6,5000	8	0,3462
1483,3 - 1698,2	6,5000	5	0,3462
1698,2 - 1913,1	6,5000	6	0,0385
1913,1 - 2180,8	6,5000	9	0,9615
$> 2180,8$	6,5000	4	0,9615
TOTAL	39	39	2,6923

Exercício 7

Relativamente às precipitações anuais em Castro D'Aire, apresentadas no Quadro A1.2, determinar:

- A precipitação associada a um período de retorno de 100 anos;
- O período de retorno do maior valor de precipitação.

Resolução:

Depois de se ter verificado (Exercício 6) que as precipitações anuais em Castro D'Aire seguem a distribuição normal é possível efetuar a análise frequencial pretendida. Para esta a análise pode recorrer-se ao **posicionamento gráfico dos dados** ou utilizar técnicas analíticas baseadas em **fatores de frequência**.

a) Utilizando a Equação (A1.40) é possível determinar a probabilidade de não excedência correspondente a $T = 100$,

$$T = \frac{1}{1 - F(x)} \Leftrightarrow F(x) = 1 - \frac{1}{T} = 1 - \frac{1}{100} = 0,99,$$

Com este valor é possível tirar da reta teórica normal, o correspondente valor de X . Para $F(x) = 99\%$ vem que $x \approx 2860$ mm (ver Figura A1.8). Isto é a **precipitação associada a um período de retorno de 100 anos é 2860 mm**.

Podemos chegar ao mesmo resultado fazendo a análise baseada em fatores de frequência. Combinando as Equações (A1.42) e (A1.44) vem,

$$x_T = \bar{x} + zS, \quad (\text{A1.48})$$

que para a distribuição em estudo é,

$$x_T = 1698,2 + z \times 498,9. \quad (\text{A1.49})$$

Já vimos que a $T = 100$ corresponde $F(x) = 0,99$ logo, da Tabela 8.1, retiramos valor de $z = 2,33$ donde da Equação (A1.49) vem,

$$x_T = 1698,2 + 2,33 \times 498,9 = 2861 \text{ mm}.$$

Isto é, a **precipitação associada a um $T = 100$ anos é 2861 mm**, que é aproximadamente igual ao valor encontrado utilizando o método do posicionamento gráfico.

b) O **maior valor de precipitação anual** em Castro D'Aire é $x = 3249,6$ mm (ver Quadro A1.2). Com este valor pode ler-se na reta teórica o correspondente valor de $F(x)$. Pela leitura da Figura A1.8. vem, para $x = 3249,6$ mm um valor de $F(x) \approx 99,91\%$. Donde,

$$T = \frac{1}{1 - F(x)} = \frac{1}{1 - 0,9991} = 1111 \text{ anos}.$$

Isto é, o **período de retorno de um valor de precipitação 3249,6 mm** é aproximadamente 1111 anos.

Para efetuar a análise baseada em fatores de frequência, utilizamos a Equação (A1.48) para encontrar o valor de z , sabendo que $x = 3249,6$,

$$3249,6 = 1698,2 + z \times 498,9 \Leftrightarrow z = \frac{3249,6 - 1698,2}{498,9} = 3,11.$$

Pela Tabela 8.1, para $z = 3,11$ vem $F(z) = 0,9991$, e o **período de retorno do valor 3249,6 mm** é,

$$T = \frac{1}{1 - F(x)} = \frac{1}{1 - 0,9991} = 1111 \text{ anos.}$$

8. Tabelas

8.1. Tabela da função de distribuição normal, $F(z)$

Que dá os valores da função de distribuição

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{z^2}{2}} dz$$

Esta tabela dá diretamente as áreas sob a curva normal compreendidas entre $z = -\infty$ e qualquer outro valor positivo de z . Para valores de $z < 0$, isto é, para valores de $z = -z$, vem que $F(-z) = 1 - F(z)$.

Tabela 8.1. Áreas referentes à curva normal reduzida, $F(z)$

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

8.2. Tabela da função de distribuição do χ^2

Tabela 8.2. Valores da estatística χ^2 , para diversos níveis de confiança $1-\alpha$ e ν graus de liberdade

ν	$1-\alpha$										
	$\chi^2_{0.995}$	$\chi^2_{0.99}$	$\chi^2_{0.975}$	$\chi^2_{0.95}$	$\chi^2_{0.90}$	$\chi^2_{0.75}$	$\chi^2_{0.50}$	$\chi^2_{0.05}$	$\chi^2_{0.025}$	$\chi^2_{0.01}$	$\chi^2_{0.005}$
1	7,88	6,63	5,02	3,84	2,71	1,32	,455	,0039	,0010	,0002	,000039
2	10,6	9,21	7,38	5,99	4,61	2,77	1,39	,103	,0506	,0201	,0100
3	12,8	11,3	9,35	7,81	6,25	4,11	2,37	,352	,216	,115	,072
4	14,9	13,3	11,1	9,49	7,78	5,39	3,36	,711	,484	,297	,207
5	16,7	15,1	12,8	11,1	9,24	6,63	4,35	1,15	,831	,554	,412
6	18,5	16,8	14,4	12,6	10,6	7,84	5,35	1,64	1,24	,872	,676
7	20,3	18,5	16,0	14,1	12,0	9,04	6,35	2,17	1,69	1,24	,989
8	22,0	20,1	17,5	15,5	13,4	10,2	7,34	2,73	2,18	1,65	1,34
9	23,6	21,7	19,0	16,9	14,7	11,4	8,34	3,33	2,70	2,09	1,74
10	25,3	23,2	20,5	18,3	16,0	12,5	9,34	3,94	3,25	2,56	2,16
11	26,8	24,7	21,9	19,7	17,3	13,7	10,3	4,57	3,82	3,05	2,60
12	28,3	26,2	23,3	21,0	18,5	14,8	11,3	5,23	4,40	3,57	3,07
13	29,8	27,7	24,7	22,4	19,8	16,0	12,3	5,89	5,01	4,11	3,57
14	31,3	29,1	26,1	23,7	21,1	17,1	13,3	6,57	5,63	4,66	4,07
15	32,8	30,6	27,5	25,0	22,3	18,2	14,3	7,26	6,26	5,23	4,60
16	34,3	32,0	28,8	26,3	23,5	19,4	15,3	7,96	6,91	5,81	5,14
17	35,7	33,4	30,2	27,6	24,8	20,5	16,3	8,67	7,56	6,41	5,70
18	37,2	34,8	31,5	28,9	26,0	21,6	17,3	9,39	8,23	7,01	6,26
19	38,6	36,2	32,9	30,1	27,2	22,7	18,3	10,1	8,91	7,63	6,84
20	40,0	37,6	34,2	31,4	28,4	23,8	19,3	10,9	9,59	8,26	7,43
21	41,4	38,9	35,5	32,7	29,6	24,9	20,3	11,6	10,3	8,90	8,03
22	42,8	40,3	36,8	33,9	30,8	26,0	21,3	12,3	11,0	9,54	8,64
23	44,2	41,6	38,1	35,2	32,0	27,1	22,3	13,1	11,7	10,2	9,26
24	45,6	43,0	39,4	36,4	33,2	28,2	23,3	13,8	12,4	10,9	9,89
25	46,9	44,3	40,6	37,7	34,4	29,3	24,3	14,6	13,1	11,5	10,5
26	48,3	45,6	41,9	38,9	35,6	30,4	25,3	15,4	13,8	12,2	11,2
27	49,6	47,0	43,2	40,1	36,7	31,5	26,3	16,2	14,6	12,9	11,8
28	51,0	48,3	44,5	41,3	37,9	32,6	27,3	16,9	15,3	13,6	12,5
29	52,3	49,6	45,7	42,6	39,1	33,7	28,3	17,7	16,0	14,3	13,1
30	53,7	50,9	47,0	43,8	40,3	34,8	29,3	18,5	16,8	15,0	13,8
40	66,8	63,7	59,3	55,8	51,8	45,6	39,3	26,5	24,4	22,2	20,7
50	79,5	76,2	71,4	67,5	63,2	56,3	49,3	34,8	32,4	29,7	28,0
60	92,0	88,4	83,3	79,1	74,4	67,0	59,3	43,2	40,5	37,5	35,5
70	104,2	100,4	95,0	90,5	85,5	77,6	69,3	51,7	48,8	45,4	43,3
80	116,3	112,3	106,6	101,9	96,6	88,1	79,3	60,4	57,2	53,5	51,2
90	128,3	124,1	118,1	113,1	107,6	98,6	89,3	69,1	65,6	61,8	59,2
100	140,2	135,8	129,6	124,3	118,5	109,1	99,3	77,9	74,2	70,1	67,3

9. Referências bibliográficas

- Chow Ven Te; Maidment D. R.; Mays, L. W. (1988). *Applied Hydrology*, McGraw-Hill, ISBN 978-0071001748, New York.
- Hipólito J. R. e Vaz, A. C. (2011). *Hidrologia e Recursos Hídricos*, IST Press, ISBN 978-9728469863, Lisboa.
- Lencastre A. e Franco F. M. (2003). *Lições de Hidrologia*, Fundação Armando Lencastre, ISBN 972-8152-59-0, Lisboa.
- Mello, F. M. (1985). *Curso de Hidrologia Aplicado à Região do Algarve. Estatística Aplicada à Hidrologia*. Universidade de Évora.