Skew t type 4 distribution – ST4

Moisés Sales

Sep 2025

Sumário

Distribuição ST4

- É uma distribuição com formato "emendado", denotada por $\texttt{ST4}(\mu, \sigma, \nu, \tau)$;

Distribuição ST4

- É uma distribuição com formato "emendado", denotada por $\mathtt{ST4}(\mu, \sigma, \nu, \tau)$;
- Ela pertence a uma família de distribuições de probabilidade que incorpora assimetria com caudas pesadas, características que a distribuição normal não possui;

Distribuição ST4

- É uma distribuição com formato "emendado", denotada por $\texttt{ST4}(\mu, \sigma, \nu, \tau)$;
- Ela pertence a uma família de distribuições de probabilidade que incorpora assimetria com caudas pesadas, características que a distribuição normal não possui;
- Usualmente são utilizadas para modelar dados que exibem assimetria e leptocurtose;

Distribuição `ST4`

- É uma distribuição com formato "emendado", denotada por $\texttt{ST4}(\mu, \sigma, \nu, \tau)$;
- Ela pertence a uma família de distribuições de probabilidade que incorpora assimetria com caudas pesadas, características que a distribuição normal não possui;
- Usualmente são utilizadas para modelar dados que exibem assimetria e leptocurtose;
- A distribuição é formalmente introduzida e descrita dentro da estrutura do pacote `GAMLSS` (Rigby et al. 2019).

Distribuição ST4

- As distribuições pertencentes à família skew-t são amplamente utilizadas na modelagem de retornos de ativos financeiros e na análise de risco, visto que a suposição de normalidade é usualmente violada e dados dessa natureza tendem a apresentar assimetria e caudas pesadas.

Distribuição `ST4`

- As distribuições pertencentes à família skew-t são amplamente utilizadas na modelagem de retornos de ativos financeiros e na análise de risco, visto que a suposição de normalidade é usualmente violada e dados dessa natureza tendem a apresentar assimetria e caudas pesadas.
- Estudos sobre curvas de crescimento de animais também se beneficiam destas distribuições; um estudo realizado por Campos e Andrade Filho (2010), utilizando `NLGAMLSS` e diferentes distribuições para o erro, incluindo distribuições skew-t, conclui que o modelo assumindo erros skew-t foi o mais versátil.

Sumário

Função Densidade de Probabilidade

Se $Y \sim \texttt{ST4}(\mu, \sigma, \nu, \tau)$, então sua função densidade de probabilidade é dada por

$$f_Y(y|\mu,\sigma,\nu,\tau) = \begin{cases} \frac{c}{\sigma}\left[1+\frac{z^2}{\nu}\right]^{-(\nu+1)/2} & \text{se} \quad y < \mu \\ \frac{c}{\sigma}\left[1+\frac{z^2}{\tau}\right]^{-(\tau+1)/2} & \text{se} \quad y \geq \mu \end{cases}$$

Função Densidade de Probabilidade

Se $Y \sim \texttt{ST4}(\mu, \sigma, \nu, \tau)$, então sua função densidade de probabilidade é dada por

$$f_Y(y|\mu,\sigma,\nu,\tau) = \begin{cases} \frac{c}{\sigma}\left[1+\frac{z^2}{\nu}\right]^{-(\nu+1)/2} & \text{se} \quad y < \mu \\ \frac{c}{\sigma}\left[1+\frac{z^2}{\tau}\right]^{-(\tau+1)/2} & \text{se} \quad y \geq \mu \end{cases}$$

em que

$$Z = \frac{Y-\mu}{\sigma},$$
$$c = 2[\sqrt{\nu}B(1/2,\nu/2) + \sqrt{\tau}B(1/2,\tau/2)]^{-1}.$$

Suporte

Intervalos		
Y	$-\infty < y < \infty$	
μ	$-\infty < \mu < \infty$	moda, parâmetro de mudança de locação
σ	$0 < \sigma < \infty$	parâmetro de escala
ν	$0 < \nu < \infty$	parâmetro de peso da cauda esquerda
τ	$0 < \tau < \infty$	parâmetro de peso da cauda direita

Tabela 1: Suporte da variável Y e dos parâmetros da distribuição ST4.

Medidas da Distribuição

- Média:
$$\mathbb{E}(Y) = \mu + \sigma\,\mathbb{E}(Z) = \mu + \sigma c\left[\frac{\tau}{(\tau-1)} - \frac{\nu}{(\nu-1)}\right]$$
para $\nu > 1$ e $\tau > 1$.

Medidas da Distribuição

- Média:

$$\mathbb{E}(Y) = \mu + \sigma\,\mathbb{E}(Z) = \mu + \sigma c \left[\frac{\tau}{(\tau - 1)} - \frac{\nu}{(\nu - 1)} \right]$$

  para $\nu > 1$ e $\tau > 1$.

- Mediana:

$$\begin{cases} \mu + \sigma t_{\frac{(1+k)}{4}\nu} & k \leq 1 \\ \mu + \sigma t_{\frac{(3k-1)}{4k}\tau} & k > 1 \end{cases}$$

Medidas da Distribuição

- Média:

$$\mathbb{E}(Y) = \mu + \sigma\,\mathbb{E}(Z) = \mu + \sigma c \left[\frac{\tau}{(\tau - 1)} - \frac{\nu}{(\nu - 1)}\right]$$

  para $\nu > 1$ e $\tau > 1$.

- Mediana:

$$\begin{cases} \mu + \sigma t_{\frac{(1+k)}{4}} \nu & k \leq 1 \\ \mu + \sigma t_{\frac{(3k-1)}{4k}} \tau & k > 1 \end{cases}$$

  em que

$$k = \frac{\sqrt{\tau} B(1/2, \tau/2)}{\sqrt{\nu} B(1/2, \nu/2)}.$$

Medidas da Distribuição

- Moda: μ
- Variância:
$$\mathrm{Var}(Y) = \sigma^2 \mathrm{Var}(Z) = \sigma^2 \left\{ \mathbb{E}(Z^2) - [\mathbb{E}(Z)]^2 \right\}$$

Medidas da Distribuição

- Moda: μ
- Variância:

$$\mathrm{Var}(Y) = \sigma^2 \mathrm{Var}(Z) = \sigma^2 \left\{ \mathbb{E}(Z^2) - [\mathbb{E}(Z)]^2 \right\}$$

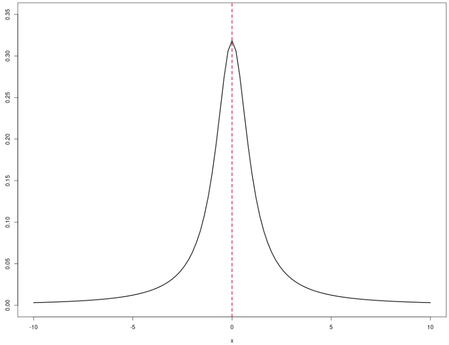
em que

$$\mathbb{E}(Z^2) = \frac{c\tau^{3/2} B(1/2, \tau/2)}{2(\tau - 2)} + \frac{c\nu^{3/2} B(1/2, \nu/2)}{2(\nu - 2)}$$

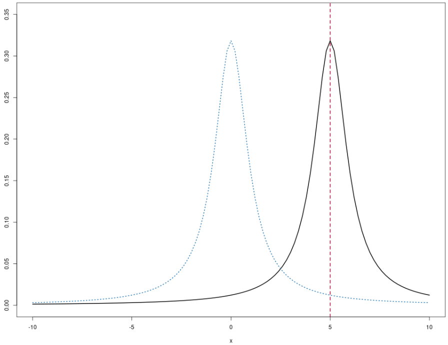
para $\nu > 2$ e $\tau > 2$.

Sumário

Variando μ



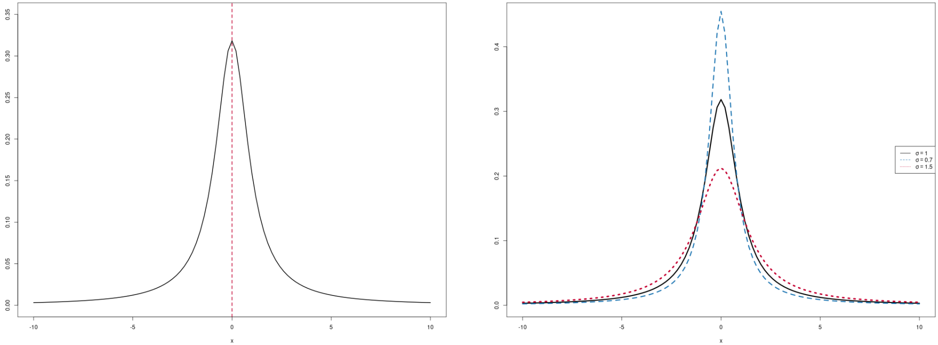
a $\mu = 0$



b $\mu = 5$

Figura 1: Forma da distribuição para diferentes valores de μ.

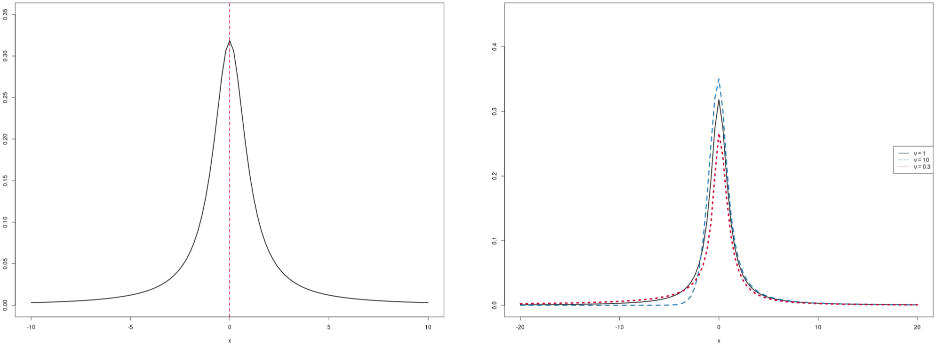
Variando σ



a $\sigma = 1$ b $\sigma = 0.7,\ 1.5$

Figura 2: Forma da distribuição para diferentes valores de σ.

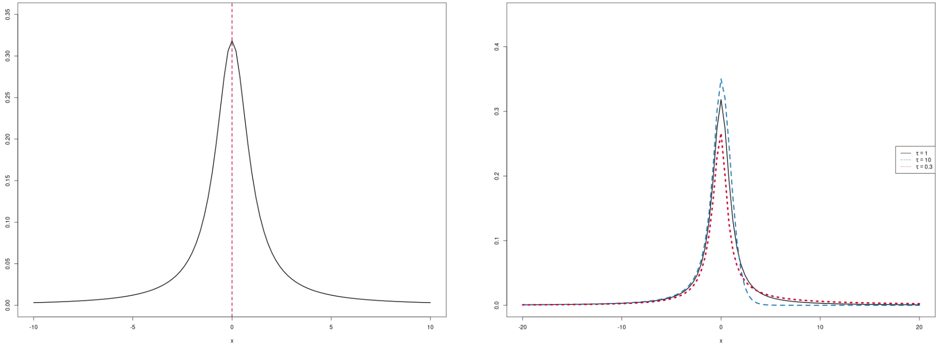
Variando ν



a $\nu = 1$ b $\nu = 0.3,\ 30$

Figura 3: Forma da distribuição para diferentes valores de ν.

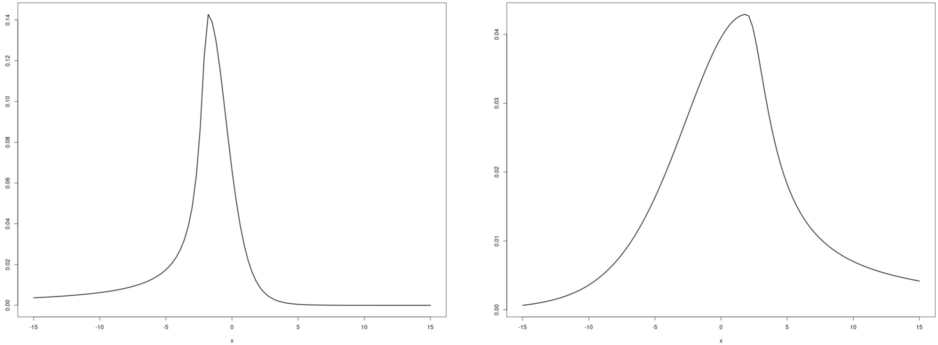
Variando τ



a $\tau = 1$ b $\tau = 0.1,\ 50$

Figura 4: Forma da distribuição para diferentes valores de τ.

Exemplos

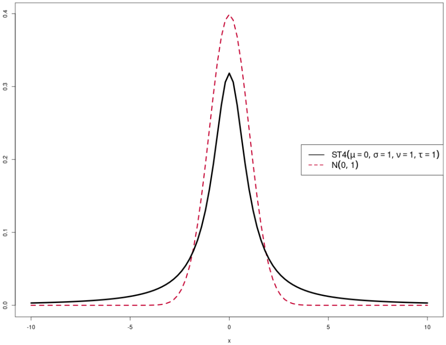


a $\mu = -1.83, \sigma = 1.5, \nu = 0.1, \tau = 7.8$

b $\mu = 1.94, \sigma = 5, \nu = 10, \tau = 0.1$

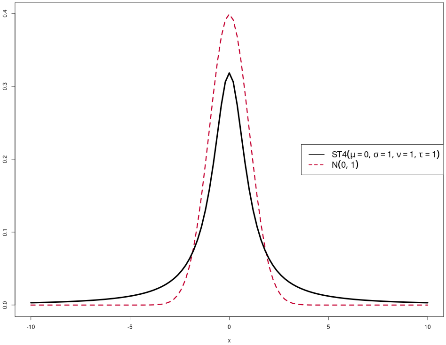
Figura 5: Formas da distribuição para diferentes valores dos parâmetros.

Comparando com distribuições mais conhecidas

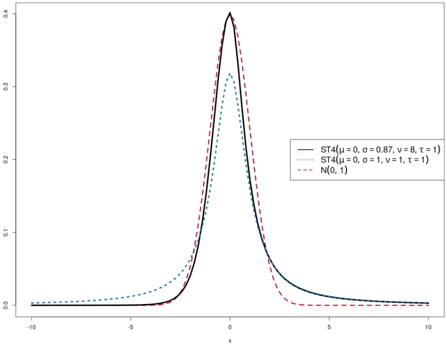


a

Comparando com distribuições mais conhecidas



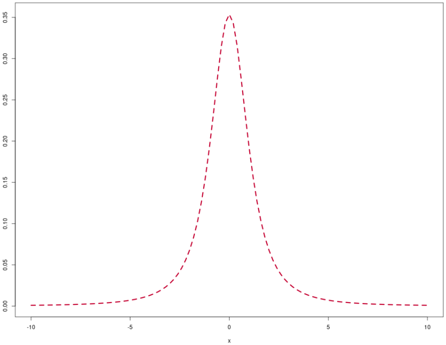
a



b

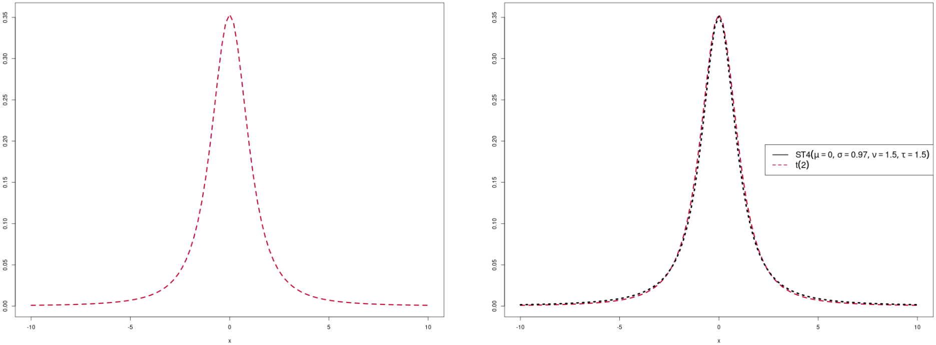
Figura 6

Comparando com distribuições mais conhecidas



a

Comparando com distribuições mais conhecidas



a b

Figura 7

Sumário

Exemplo Prático – Metodologia

- Além do estudo "Ajuste de curvas de crescimento usando NLGAMLSS" de Campos e Andrade Filho (2010) previamente apresentado, mais uma aplicação na área da biologia é dada por "Modelagem Bayesiana para curvas de crescimentos de codornas assumindo assimetria nos erros" de Rossi e Santos (2014);

Exemplo Prático – Metodologia

- Além do estudo "Ajuste de curvas de crescimento usando NLGAMLSS" de Campos e Andrade Filho (2010) previamente apresentado, mais uma aplicação na área da biologia é dada por "Modelagem Bayesiana para curvas de crescimentos de codornas assumindo assimetria nos erros" de Rossi e Santos (2014);
- O objetivo do trabalho foi avaliar o ajuste de diferentes modelos não-lineares a dados de peso corporal de codornas assumindo diferentes distribuições para o erro sob o ponto de vista Bayesiano;

Exemplo Prático – Metodologia

- O experimento utilizou 1.831 codornas de corte (*Coturnix coturnix japonica*), sendo constituido de 903 fêmeas e 928 machos.

Exemplo Prático – Metodologia

- O experimento utilizou 1.831 codornas de corte (*Coturnix coturnix japonica*), sendo constituido de 903 fêmeas e 928 machos.
- Os animais foram pesados semanalmente, formando um banco de dados para peso corporal (em gramas) ao nascimento, 7, 14, 21, 28 e 35 dias de idade.

Exemplo Prático – Metodologia

- O estudo foi constituido das seguintes etapas:
    1. Consistiu em considerar alguns modelos não-lineares mais utilizados para descrever curvas de crescimento de aves em diferentes formas parametrizadas, assumindo-se distribuição normal para os erros e selecionar aquele que melhor se ajusta aos dados por meio da inferência Bayesiana.

Exemplo Prático – Metodologia

- O estudo foi constituido das seguintes etapas:
    1. Consistiu em considerar alguns modelos não-lineares mais utilizados para descrever curvas de crescimento de aves em diferentes formas parametrizadas, assumindo-se distribuição normal para os erros e selecionar aquele que melhor se ajusta aos dados por meio da inferência Bayesiana.
    2. Consideram-se distribuições alternativas à normal para o erro, sendo elas a distribuição t, a skew-normal e skew-t.

Exemplo Prático – Resultados

- Os autores concluem tanto em um estudo simulado, quanto em uma aplicação a dados reais, que ao utilizar a curva de crescimento descrita por

  $$f(x) = \beta_1(1 - \beta_2 e^{\beta_3 x})^3$$

  o modelo skew-t para os erros é o mais adequado por apresentar maior probabilidade de cobertura dos intervalos de credibilidade para os verdadeiros valores dos parâmetros e menor DIC (*Deviance Information Criterion*).

Exemplo Prático – Resultados

Idade (dias)	Pesos (g)			
	Machos		Fêmeas	
	Média (dp)	$P_{2.5\%} - P_{97.5\%}$	Média (dp)	$P_{2.5\%} - P_{97.5\%}$
1	$9{,}50^a$ (0,031)	9,45 - 9,57	$9{,}58^a$ (0,033)	9,52 - 9,65
7	$29{,}66^a$ (0,164)	29,34 - 29,98	$29{,}81^a$ (0,169)	29,47 - 30,14
14	$76{,}56^a$ (0,314)	75,95 - 77,18	$76{,}88^a$ (0,365)	76,17 - 77,60
21	$139{,}00^b$ (0,472)	138,10 - 139,90	$145{,}50^a$ (0,592)	140,40 - 142,70
28	$192{,}30^b$ (0,553)	191,20 - 193,40	$197{,}50^a$ (0,656)	196,20 - 198,80
35	$225{,}70^b$ (0,639)	224,40 - 226,90	$235{,}90^a$ (0,724)	234,50 - 237,40

dp= desvio-padrão da média; $P_{2.5\%} - P_{97.5\%}$ = *intervalo com 95% de credibilidade;*

^{a,b} *Letras distintas na linha, indicam diferenças significativas entre as médias dos pesos (g), por meio de comparações Bayesianas em nível de 95% de credibilidade.*

Fonte: Rossi e Santos (2014).

Exemplo Prático – Resultados

A segunda etapa da análise indicou que assumindo erros skew-normais e skew-t para os machos e fêmeas, respectivamente, obtemos os melhores ajustes para os dados de pesos corporais de codornas.

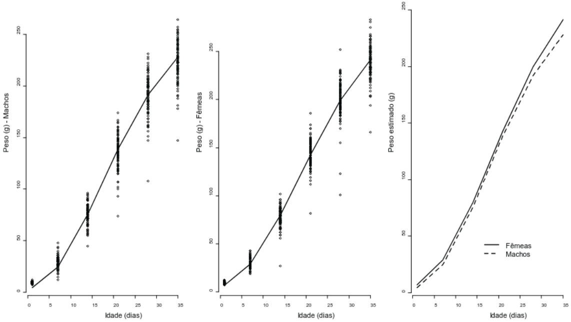


Figura 8: Curvas de *Gompertz* - f_2 ajustadas a dados de peso (g), por meio do procedimento P_1, com erros skew-normais e skew-t, respectivamente, para machos e fêmeas.

Fonte: Rossi e Santos (2014).

Exemplo Prático – Conclusões

O modelo não-linear de Gomperts na forma parametrizada f_2, com erros skew-normais e skew-t, respectivamente, para machos e fêmeas, por meio do procedimento proposto por De la Cruz e Branco (2009), foi o que melor ajustou os dados de pesos corporal das codornas.

Sumário

Conclusão

- A distribuição pode ser utilizada quando a distribuição dos dados assume valores na reta real e possui natureza assimétrica com caudas pesadas.

Conclusão

- A distribuição pode ser utilizada quando a distribuição dos dados assume valores na reta real e possui natureza assimétrica com caudas pesadas.
- As principais vantages se resuem a grande capacidade de adaptabilidade da distribuição, podemos alterar o peso de cada uma de suas caudas, podemos alterar a sua curtose e sua locação, viabilizando uma maior aderência aos dados.

Conclusão

- A distribuição pode ser utilizada quando a distribuição dos dados assume valores na reta real e possui natureza assimétrica com caudas pesadas.
- As principais vantages se resuem a grande capacidade de adaptabilidade da distribuição, podemos alterar o peso de cada uma de suas caudas, podemos alterar a sua curtose e sua locação, viabilizando uma maior aderência aos dados.
- A distribuição é limitada quando se tratam de dados que assumem valores somente positivos ou negativos.

Conclusão

- A distribuição pode ser utilizada quando a distribuição dos dados assume valores na reta real e possui natureza assimétrica com caudas pesadas.
- As principais vantages se resuem a grande capacidade de adaptabilidade da distribuição, podemos alterar o peso de cada uma de suas caudas, podemos alterar a sua curtose e sua locação, viabilizando uma maior aderência aos dados.
- A distribuição é limitada quando se tratam de dados que assumem valores somente positivos ou negativos.
- A forma da distribuição pode ser mostrada graficamente por meio do comando `demo.ST4()`, em que, de forma interativa, todos os parâmetros podem ser ajustados.

Conclusão

- A distribuição pode ser utilizada quando a distribuição dos dados assume valores na reta real e possui natureza assimétrica com caudas pesadas.
- As principais vantages se resuem a grande capacidade de adaptabilidade da distribuição, podemos alterar o peso de cada uma de suas caudas, podemos alterar a sua curtose e sua locação, viabilizando uma maior aderência aos dados.
- A distribuição é limitada quando se tratam de dados que assumem valores somente positivos ou negativos.
- A forma da distribuição pode ser mostrada graficamente por meio do comando `demo.ST4()`, em que, de forma interativa, todos os parâmetros podem ser ajustados.
- Um método alternativo para mostrar a função densidade de probabilidade é utilizar a função `pdf.plot(family = ST4, mu = c(0, 3), sigma = c(1, 5), nu = c(1, 0.5), tau = 1)`

Sumário

Referências

- Campos, AM e Marinho Gomes de Andrade Filho (2010). "Ajuste de curvas de crescimento usando NLGAMLSS". Em: *19º Simpósio Nacional de Probabilidade e Estatística*, pp. 1–5.
- De la Cruz, Rolando e Márcia D Branco (2009). "Bayesian analysis for nonlinear regression model under skewed errors, with application in growth curves". Em: *Biometrical Journal: Journal of Mathematical Methods in Biosciences* 51.4, pp. 588–609.
- Rigby, Robert A et al. (2019). *Distributions for modeling location, scale, and shape: Using GAMLSS in R*. Chapman e Hall/CRC.
- Rossi, Robson Marcelo e Lucimary Afonso dos Santos (2014). "Modelagem Bayesiana para curvas de crescimentos de codornas assumindo assimetria nos erros". Em: *Semina: Ciências Agrárias* 35.3, pp. 1637–1647.