

# Distribuição Weibull Inversa Generalizada

Ada Maris Pereira Mário, Emanuel Victor da Silva Favorato

**Abstract**—Neste projeto, apresenta-se uma análise da distribuição Weibull Inversa Generalizada, seu desenvolvimento histórico, propriedades e demais aspectos teóricos, principalmente no que tange a análise de sobrevivência. Para tal, utilizou-se de ferramentas da linguagem R, especialmente a biblioteca *RelDists*, para caracterização do modelo e seu comportamento, e a biblioteca *gamlss*, para realização de um mini-estudo de simulação e respectiva estimação de seus parâmetros, bem como para aplicações práticas em dados de mortalidade por câncer e comparações com modelos mais simples e usuais. A distribuição mostra-se bastante útil para estudos industriais e epidemiológicos, porém, pode não ser a melhor escolha para os dados utilizados no contexto deste projeto, tendo maior proveito em situações que exigem modelos mais complexos.

**Index Terms**—Distribuição Weibull Inversa Generalizada, sobrevivência, *gamlss*, caracterização, comparação de modelos.

## I. ETAPA 01

### ATUALIZAÇÃO HISTÓRICA DO MODELO (ENTREGA EM 22/03/2024)

Introduzida em meados dos anos 1930, a Distribuição Weibull teve suas várias aplicabilidades enumeradas pelo estudioso de quem recebeu o nome apenas em 1951 [1]. Mais de 70 anos depois, a distribuição foi e é reconhecida pela sua ampla versatilidade em diversos problemas, em especial no que tange a análise de sobrevivência. Sua função de densidade acumulada (cdf) é dada por:

$$F(x; \lambda, k) = 1 - e^{-\left(\frac{x}{\lambda}\right)^k}; \lambda > 0; k > 0 \quad (1)$$

Por sua vez, a nomeada distribuição Weibull inversa (IW), ou até mesmo Weibull complementar [2], tem cdf dada por:

$$G(t; \lambda, k) = e^{-\left(\frac{\lambda}{t}\right)^k} \quad (2)$$

A distribuição Weibull inversa tem sido usada para analisar dados de vida útil com algumas taxas de falha monótonas. É um modelo adequado para descrever fenômenos de degradação de componentes mecânicos como pistões e virabrequins de motores diesel. Também tem sido usado para modelar a decomposição de fluido isolante [3] [4].

Já Gusmão et al (2009) [5] propôs o modelo generalizado da distribuição anteriormente citada, o qual é o tema deste artigo. Ela inclui as distribuições Weibull inversa, exponencial inversa e Rayleigh inversa

Com características favoráveis para estudos biológicos e de confiabilidade [6], a distribuição GIW tem sido aplicada em uma ampla gama de situações, incluindo aplicações em medicina, confiabilidade e ecologia. Ela também pode ser usada para descrever o fenômeno de degradação de componentes mecânicos [7]. Durante a pandemia do coronavírus, o modelo

foi bastante utilizado para modelagem de casos e mortes por COVID-19 [8] [9] [10].

Em [8], os autores aplicaram um modelo baseado em aprendizado de máquina para prever a ameaça potencial da COVID-19 em países ao redor do mundo. Eles mostraram que, usando ponderação iterativa para ajustar a GIW, pode-se obter um ajuste melhor para desenvolver uma estrutura de previsão. A distribuição apresenta um desempenho melhor que a distribuição normal para ajuste dos dados e é chamada de "Weibull robusta" na pesquisa.

Muitas modificações da Weibull similares à GIW foram propostas nos últimos anos. Inspirado pela noção da família gama-generalizada de distribuições de Zografos e Balakrishnan, a classe de distribuições Weibull generalizada foi proposta [11]. Outro exemplo são os estudos emergentes da distribuição beta Weibull generalizada [12]. Além disso, uma variedade de modelos tem sido obtida na tentativa de modelar outras formas da função taxa de falha [13].

Por outro lado, Almalki e Nadarajah [14] afirmam que a proposta de Gusmão et al. é apenas uma simples reparametrização da distribuição Weibull inversa, e não uma generalização.

## II. ETAPA 02A E 02B

### CARACTERIZAÇÃO DO MODELO (ENTREGA EM 19/04/2024)

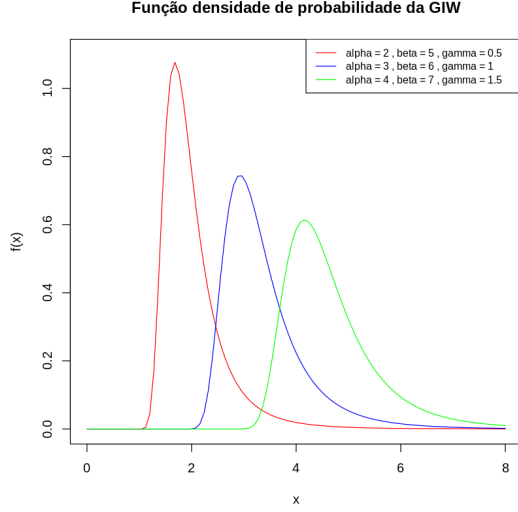
A distribuição tri-paramétrica possui função de densidade de probabilidade dada por:

$$f(t) = \gamma \beta \alpha^\beta t^{-(\beta+1)} e^{-\gamma \left(\frac{\alpha}{t}\right)^\beta}, t > 0 \quad (3)$$

Onde  $\alpha, \beta, \gamma > 0$ .

Gusmão et al. mostrou que a distribuição IW é um caso particular da generalizada (GIW) quando  $\gamma = 1$  em (3). Tal função de densidade é provada unimodal em [6].

Podemos verificar sua curva de densidade com parâmetros arbitrários e variados utilizando a biblioteca *RelDists* em R. Abaixo temos um exemplo com os vetores de parâmetros  $\alpha$  (ou  $\mu$ , segundo a função da biblioteca) =  $\{2, 3, 4\}$ , parâmetro de locação;  $\beta$  (ou  $\sigma$ ) =  $\{5, 6, 7\}$ , parâmetro de forma; e  $\gamma$  (ou  $\nu$ ) =  $\{0.5, 1, 1.5\}$ , parâmetro que tem por finalidade aumentar a flexibilidade da distribuição [6].

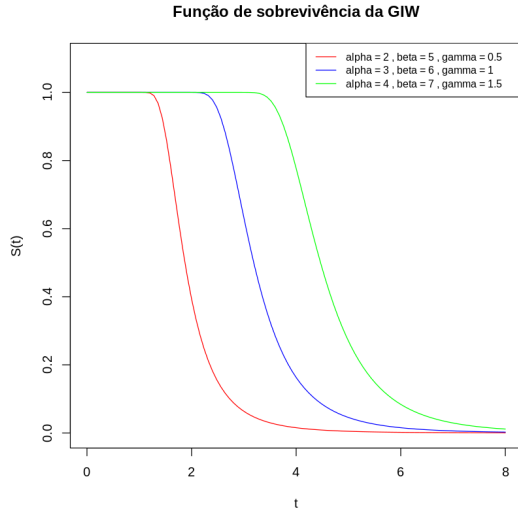


**Fig. 1:** Gráfico da função de densidade com diferentes parâmetros

A função de sobrevivência para tal distribuição é dada por:

$$S(t) = 1 - F(t) = 1 - e^{-\gamma(\frac{\alpha}{t})^\beta} \quad (4)$$

Abaixo está o plot utilizando os mesmos vetores de parâmetros especificados anteriormente:

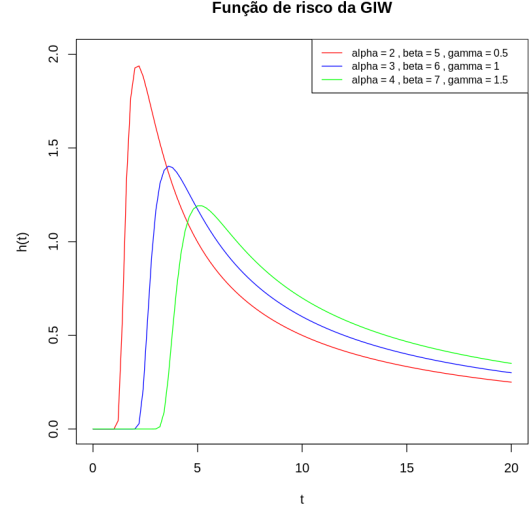


**Fig. 2:** Gráfico da função de sobrevivência com diferentes parâmetros

Já sua função de risco (ou taxa de falha) é:

$$h(t) = \gamma\beta\alpha^\beta t^{-(\beta+1)} e^{-\gamma(\frac{\alpha}{t})^\beta} [1 - e^{-\gamma(\frac{\alpha}{t})^\beta}]^{-1} \quad (5)$$

Abaixo está o plot desta função:



**Fig. 3:** Gráfico da função de risco com diferentes parâmetros

No artigo de referência deste estudo, o autor mostra que  $h(t)$  é unimodal, o que pode ser observado no próprio gráfico. A partir da função  $h_{GIW}$  da biblioteca *RelDists*, verificou-se que os pontos no tempo em que o risco é máximo, para os parâmetros antes estabelecidos, foram  $t = \{2.12, 3.63, 5.10\}$ .

A partir da função geradora de momentos dada pela expressão abaixo, conseguimos obter a esperança dos  $T^k$  momentos populacionais:

$$M(z) = E(e^{zT}) = \sum_{k=0}^n \frac{z^k}{k!} \gamma^{\frac{k}{\beta}} \alpha^k \Gamma(1 - k\beta^{-1}) \quad (6)$$

$$E(T^k) = \gamma^{\frac{k}{\beta}} \alpha^k \Gamma(1 - k\beta^{-1}) \quad (7)$$

Podemos, assim, obter as expressões da média ( $E(T)$ ) e variância ( $Var(T)$ ) utilizando as expressões acima:

$$E(T) = \gamma^{\frac{1}{\beta}} \alpha \Gamma(1 - \beta^{-1}) \quad (8)$$

$$\begin{aligned} Var(T) &= E(T^2) - [E(T)]^2 \\ &= \gamma^{\frac{2}{\beta}} \alpha^2 \Gamma(1 - 2\beta^{-1}) - (\gamma^{\frac{1}{\beta}} \alpha \Gamma(1 - \beta^{-1}))^2 \\ &= \gamma^{\frac{2}{\beta}} \alpha^2 (\Gamma(1 - 2\beta^{-1}) - \Gamma^2(1 - \beta^{-1})) \end{aligned} \quad (9)$$

A moda da distribuição pode ser obtida derivando a distribuição em relação a variável aleatória  $T$ ; logo, temos que sua expressão é dada por:

$$\begin{aligned}
M(T) &= \frac{d}{dT}(\gamma\beta\alpha^\beta t^{-(\beta+1)} e^{-\gamma(\frac{\alpha}{t})^\beta}) \\
&= \gamma\beta\alpha^\beta \frac{d}{dT}(t^{-(\beta+1)} e^{-\gamma(\frac{\alpha}{t})^\beta}) \\
&= \gamma\beta\alpha^\beta (e^{-\gamma(\frac{\alpha}{t})^\beta} \frac{d}{dT}(t^{-(\beta+1)}) + t^{-(\beta+1)} \frac{d}{dT}(e^{-\gamma(\frac{\alpha}{t})^\beta})) \\
&= \gamma\beta\alpha^\beta (e^{-\gamma(\frac{\alpha}{t})^\beta} (-(\beta+1)t^{-(\beta+2)}) + \\
&\quad t^{-(\beta+1)} e^{-\gamma(\frac{\alpha}{t})^\beta} \frac{d}{dT}(-\gamma(\frac{\alpha}{t})^\beta)) \\
&= \gamma\beta\alpha^\beta (e^{-\gamma(\frac{\alpha}{t})^\beta} (-(\beta+1)t^{-(\beta+2)} + t^{-(\beta+1)} \gamma\beta \frac{\alpha^\beta}{t^{\beta+1}})) \\
&= \gamma\beta\alpha^\beta e^{-\gamma(\frac{\alpha}{t})^\beta} t^{-(\beta+1)} (\frac{-(\beta+1)}{t} + \gamma\beta \frac{\alpha^\beta}{t^{\beta+1}})
\end{aligned} \tag{10}$$

Igualando  $M(T)$  a zero, encontraremos o  $t$  de maior probabilidade de ocorrência, isto é, o valor numérico da moda, dado pela expressão:

$$\begin{aligned}
\frac{-(\beta+1)}{t} + \gamma\beta \frac{\alpha^\beta}{t^{\beta+1}} &= 0 \\
\frac{-(\beta+1)t^\beta + \gamma\beta\alpha^\beta}{t^{\beta+1}} &= 0 \\
(\beta+1)t^\beta &= \gamma\beta\alpha^\beta \\
t &= \left(\frac{\gamma\beta\alpha^\beta}{\beta+1}\right)^{\frac{1}{\beta}}
\end{aligned} \tag{11}$$

Com os mesmos vetores de parâmetros escolhidos anteriormente, foram gerados 10000 valores aleatórios da distribuição GIW por meio da função *rGIW* da biblioteca *RelDists*. A partir de tal amostra, calculou-se sua média, variância, mediana e moda, obtendo-se, respectivamente,  $\bar{x} = \{2.03, 3.38, 4.67\}$ ,  $s^2 = \{0.42, 0.65, 0.99\}$ ,  $\hat{M}d = \{1.88, 3.19, 4.47\}$ ,  $\hat{M}o = \{1.68, 2.92, 4.16\}$ . Já os valores populacionais foram calculados substituindo-se os parâmetros nas equações encontradas acima. Desse modo, obteve-se  $\mu = \{2.03, 3.39, 4.69\}$ ,  $\sigma^2 = \{0.41, 0.72, 0.96\}$ ,  $Md = \{1.87, 3.19, 4.47\}$ ,  $Mo = \{1.68, 2.92, 4.16\}$ . Observa-se que os vetores das duas últimas medidas ficaram idênticos, enquanto que os vetores de médias e variâncias ficaram bem próximos, respaldando os cálculos realizados. Ademais, tais resultados permitem concluir que a distribuição é assimétrica à esquerda, uma vez que os valores das médias são sempre maiores que os das medianas, que por sua vez são maiores do que as modas.

Os percentis da GIW podem ser obtidos através da função acumulada inversa, de tal forma que:

$$\begin{aligned}
p &= e^{-\gamma(\frac{\alpha}{t})^\beta} \\
\ln(p) &= -\gamma(\frac{\alpha}{t})^\beta \\
\frac{-\ln(p)}{\gamma} &= (\frac{\alpha}{t})^\beta \\
(\frac{-\ln(p)}{\gamma})^{\frac{1}{\beta}} &= \frac{\alpha}{t} \\
t &= \alpha(\frac{-\gamma}{\ln(p)})^{\frac{1}{\beta}}
\end{aligned} \tag{12}$$

É importante destacar que o valor da mediana pode ser obtido substituindo-se  $p$  por 0.5 na expressão anterior.

Já para a vida média residual, temos:

$$\begin{aligned}
E(T-t | T > t) &= \frac{\int_t^\infty S(u) du}{S(t)} \\
E(T-t | T > t) &= \left[ \frac{\int_t^\infty 1 - e^{-\gamma(\frac{\alpha}{u})^\beta} du}{1 - e^{-\gamma(\frac{\alpha}{t})^\beta}} \right]
\end{aligned} \tag{13}$$

A expressão exata da integral apresentada no denominador da expressão não pode ser obtida com base nos métodos de integrabilidade tradicionais, exigindo a aplicação de métodos matemáticos mais sofisticados, cuja demonstração requer trabalhos mais complexos que não serão abordados no estudo.

### III. ETAPA 03A ESTUDO DE SIMULAÇÃO (ENTREGA EM 03/05/2024)

Utilizando a linguagem *R*, implementou-se uma função para gerar dados aleatórios de uma GIW, calcular suas estatísticas e plotar seu histograma a partir dos parâmetros  $n$ ,  $\alpha$ ,  $\beta$  e  $\gamma$ .

Na função, as estatísticas amostrais são calculadas por meio de funções da própria linguagem como *quantile* para os quantis, *median* para a mediana, *mean* para a média e *var* para a variância. Já as estatísticas populacionais são calculadas por meio da definição das equações 8 e 9, obtidas na etapa II deste projeto.

Com tais valores, são calculados os vieses (14) e os quadrados médios dos erros (15) dos parâmetros.

$$Bias(\theta, \hat{\theta}) = |\theta - \hat{\theta}| \tag{14}$$

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] \tag{15}$$

A estimação dos parâmetros a partir dos dados simulados é feita utilizando o pacote *gamlss* com função de ligação de cada parâmetro definida como *log*, uma vez que todos os parâmetros da GIW devem ser positivos.

A partir disso, é plotado um histograma dos dados simulados, a curva real da GIW com os parâmetros recebidos, representada pela curva vermelha, e a curva observada com os parâmetros estimados, curva azul.

Foi definida uma matriz de cenários a serem simulados com os seguintes vetores:  $n = \{100, 1000\}$ ;  $\alpha = \{2, 4\}$ ;  $\beta = \{5, 10\}$ ;  $\gamma = \{1, 2\}$ . Desse modo, obtiveram-se 16 simulações, as quais seus respectivos histogramas estão expressos abaixo:

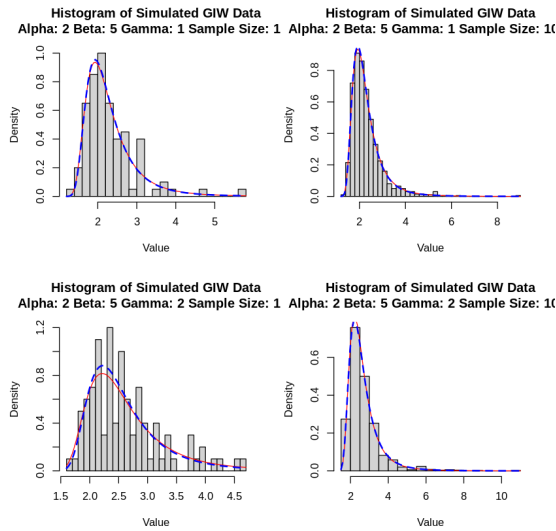


Fig. 4: Histograma dos cenários 1 - 4

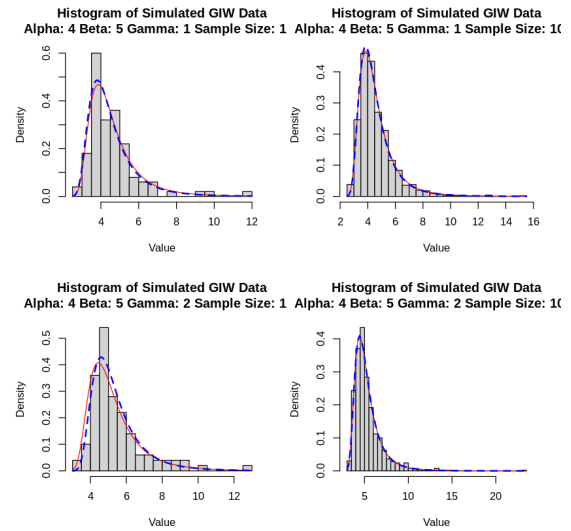


Fig. 6: Histograma dos cenários 9 - 12

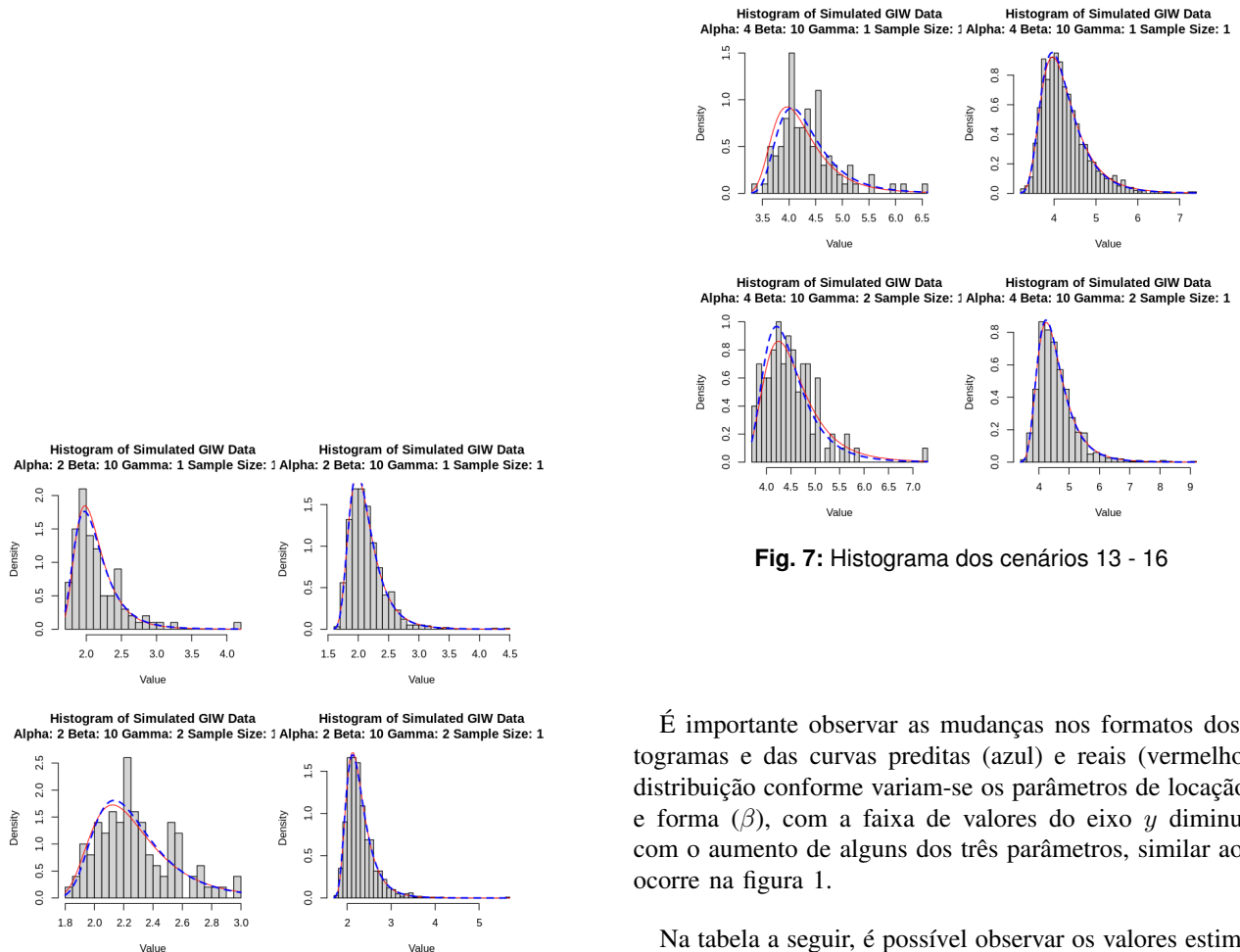


Fig. 7: Histograma dos cenários 13 - 16

É importante observar as mudanças nos formatos dos histogramas e das curvas preditas (azul) e reais (vermelho) da distribuição conforme variam-se os parâmetros de locação ( $\alpha$ ) e forma ( $\beta$ ), com a faixa de valores do eixo  $y$  diminuindo com o aumento de alguns dos três parâmetros, similar ao que ocorre na figura 1.

Na tabela a seguir, é possível observar os valores estimados para os parâmetros  $\alpha$ ,  $\beta$  e  $\gamma$ , bem como seus respectivos vieses:

n	$\alpha$	$\beta$	$\gamma$	Bias( $\alpha$ )	Bias( $\beta$ )	Bias( $\gamma$ )
100	2	5	1	0.1756768	0.09627128	0.3481061
1000	2	5	1	0.1701807	0.09885268	0.3500428
100	2	5	2	0.4850671	0.4073703	1.351729
1000	2	5	2	0.5068742	0.112954	1.346179
100	2	10	1	0.109924	0.4918808	0.4140341
1000	2	10	1	0.1970128	0.1661983	0.6156455
100	2	10	2	0.4288802	0.5435747	1.719248
1000	2	10	2	0.3133669	0.1680938	1.53079
100	4	5	1	0.2834965	0.1230807	0.3491447
1000	4	5	1	0.3247101	0.1278803	0.3483997
100	4	5	2	1.140576	0.4360049	1.356786
1000	4	5	2	1.007973	0.02659046	1.347694
100	4	10	1	0.09300058	0.1050998	0.009147592
1000	4	10	1	0.4026295	0.2974508	0.6375486
100	4	10	2	0.7683961	1.10501	1.729512
1000	4	10	2	0.6712393	0.1516354	1.59004

**TABLE I:** Tabela com os vieses para os casos simulados

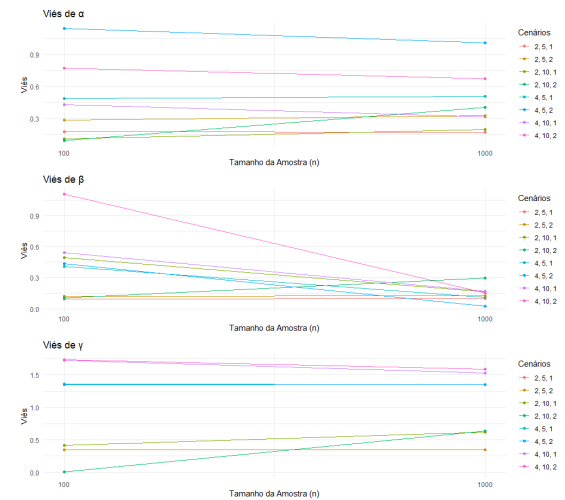
Nota-se que, para praticamente todos os cenários, os valores estimados de  $\gamma$  têm um viés muito alto em comparação aos outros parâmetros, mesmo com modificações do valor real ou do tamanho da amostra.

Já na tabela II, temos os valores observados para o erro quadrático médio (EQM) de cada parâmetro:

n	$\alpha$	$\beta$	$\gamma$	EQM( $\alpha$ )	EQM( $\beta$ )	EQM( $\gamma$ )
100	2	5	1	0.03086236	0.00926816	0.1211779
1000	2	5	1	0.02896147	0.009771853	0.12253
100	2	5	2	0.2352901	0.1659505	1.82717
1000	2	5	2	0.2569215	0.01275861	1.812199
100	2	10	1	0.01208328	0.2419468	0.1714242
1000	2	10	1	0.03881403	0.02762186	0.3790193
100	2	10	2	0.1839382	0.2954735	2.955812
1000	2	10	2	0.09819884	0.02825552	2.343317
100	4	5	1	0.08037028	0.01514885	0.121902
1000	4	5	1	0.1054367	0.01635338	0.1213823
100	4	5	2	1.300914	0.1901002	1.840868
1000	4	5	2	1.01601	0.0007070526	1.816278
100	4	10	1	0.008649108	0.01104596	0.0000836784
1000	4	10	1	0.1621105	0.088477	0.4064682
100	4	10	2	0.5904326	1.221046	2.991212
1000	4	10	2	0.4505622	0.02299329	2.528227

**TABLE II:** Tabela com o EQM para os casos simulados

Os valores relativamente baixos dos erros quadráticos médios, especialmente quando comparados com os vieses, nos permitem observar que os estimadores apresentam baixa variabilidade na predição.



**Fig. 8:** Vieses dos estimadores e para cada valor de  $n$

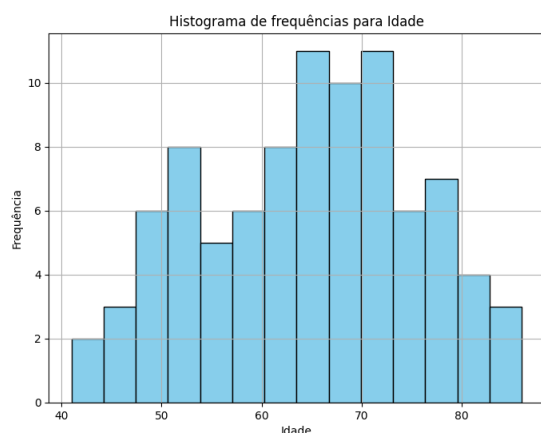
Na figura 8, têm-se os gráficos dos vieses para os estimadores dos parâmetros  $x$  tamanho da amostra para cada cenário. Observa-se que os vieses de  $\alpha$  e  $\gamma$  se mantêm praticamente os mesmos ainda que se aumente o valor de  $n$ . Já  $\beta$  tem diminuições mais abruptas conforme a amostra aumenta. Nota-se, também, o efeito contrário nos três parâmetros para os cenários iniciais, quando seus respectivos valores arbitrários são baixos. Porém, os vieses destes casos (valores pequenos dos parâmetros), são menores que os vícios de quando os parâmetros têm maiores valores.

#### IV. ETAPA 04 ESTUDO PRÁTICO (ENTREGA EM 07/06/2024)

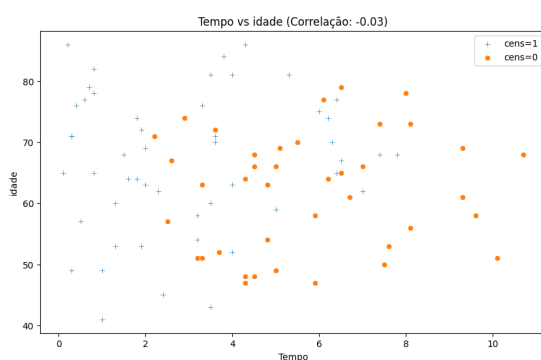
Para esta etapa, foi utilizado um conjunto de dados de um estudo sobre câncer de laringe [15]. Os dados, com 90 entradas, referem-se a um estudo realizado com 90 pacientes do sexo masculino diagnosticados no período de 1970 a 1978 com câncer de laringe e que foram acompanhados até 01/01/1983. Para cada paciente, foram registrados, no diagnóstico, a idade (em anos) e o estágio da doença (I = tumor primário, II = envolvimento de nódulos, III = metástases e IV = combinações dos 3 estágios anteriores), bem como seus respectivos tempos de morte ou censura (em meses). Os estágios se encontram ordenados pelo grau de seriedade da doença (menos sério para mais sério).

O intuito principal desta etapa é ajustar modelos de regressão paramétricos e semi-paramétricos — sendo um deles a distribuição estudada — considerando as covariáveis do conjunto de dados.

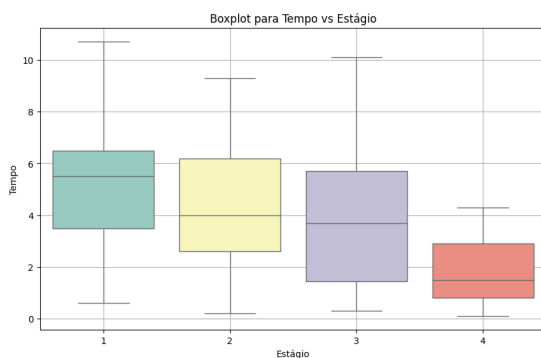
Inicialmente, realizou-se uma análise exploratória dos dados com o intuito de identificar possíveis padrões e tendências que se destacariam no estudo de sobrevivência. A seguir, apresentam-se alguns resultados e visualizações:



**Fig. 9:** Histograma para a distribuição das idades



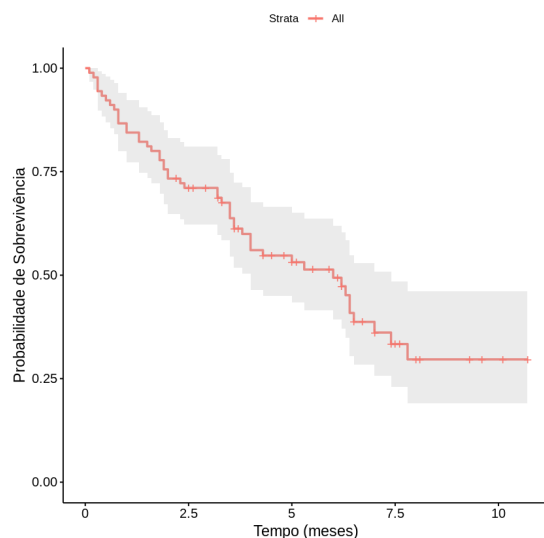
**Fig. 10:** Correlação tempo vs idade, onde cens = 0 indica óbito e cens = 1 censura



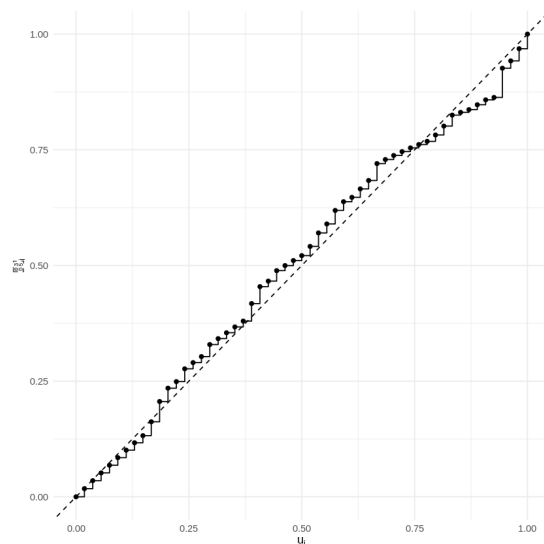
**Fig. 11:** Boxplot Tempo vs Estágio

Como pode ser observado com base nas visualizações de dados acima, temos que a figura 9 nos indica que, com base na amostra observada, tal câncer é mais comum entre idosos, indivíduos com 60 anos ou mais. Já na figura 10 observamos que a correlação entre idade e tempo para vir a óbito é negligenciável, o que indica que indivíduos mais jovens não necessariamente virão a facer mais tardiamente, por fim, a figura 11 nos mostra que, com base na análise de boxplots, não podemos concluir que haja qualquer relação entre o estágio em que a doença se encontra e o tempo para os indivíduos virem a óbito, sendo assim, observamos pouca interação real entre as variáveis observadas no estudo.

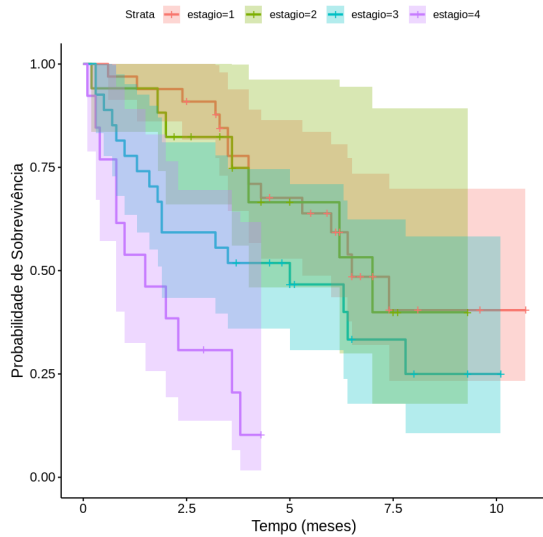
Por outro lado, pode-se observar as curvas de sobrevivência estimadas por Kaplan-Meier e seus respectivos TTTplots:



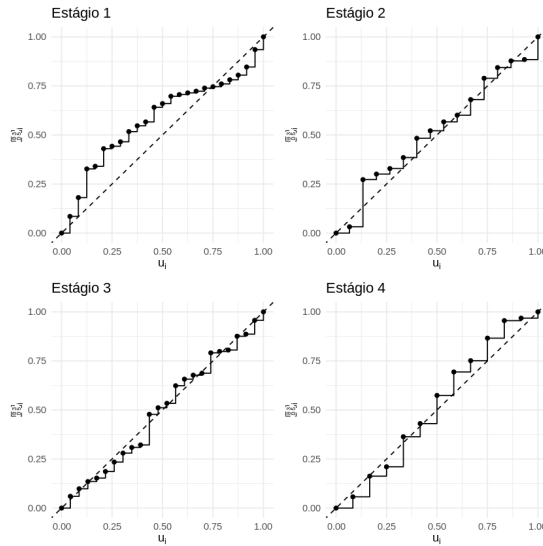
**Fig. 12:** Curva de Sobrevivência de Kaplan-Meier



**Fig. 13:** TTTplot - Dados Gerais



**Fig. 14:** Curva de Sobrevida de Kaplan-Meier Segmentada por Estágios



**Fig. 15:** TTTplots - Estágios

Pelas figuras 12 e 13, tem-se uma queda gradual ao longo do tempo, estabilizando-se por volta dos 7 meses e meio, e um risco crescente na maior parte do tempo, mas não muito distante da linha de referência. Porém, ao separar as curvas por estágio na figura 14, observam-se quedas mais acentuadas para os estágios mais avançados, mas ainda com intervalos de confiança bastante amplos, indicando a possibilidade de comportamento similar a outros estágios, sem qualquer relação com o tempo para o evento, como indicado na figura 11. Enquanto isso, os plots da figura 15 apresentam uma suposta mudança de comportamento do risco conforme os estágios avançam, com estes sempre se intensificando nos tempos finais. Entretanto, não se pode descartar a possibilidade de riscos constantes e similares para todos os estágios, visto que estão todos muito próximos da linha de referência.

Feita a análise inicial, realizaram-se ajustes para os modelos paramétricos de regressão Exponencial, Weibull e, a

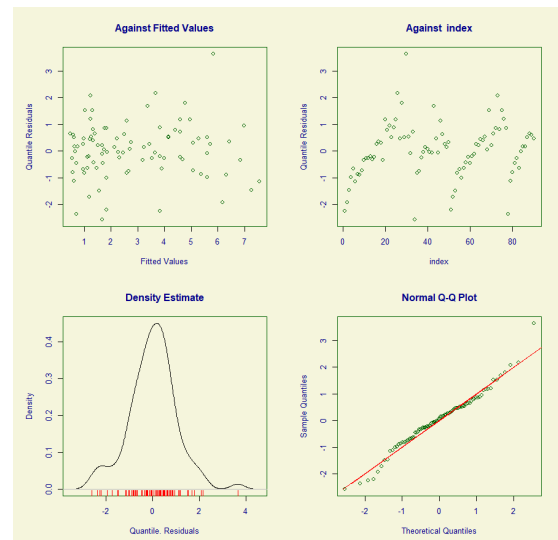
distribuição em pauta, GIW. Para tal, utilizaram-se as bibliotecas *gamlss* e *RelDists* do R. A seguir, apresenta-se um resumo das regressões utilizando as covariáveis *idade* e *estagio* e censuras à direita a partir dos dados.

**TABLE III:** Resumo dos Coeficientes de Regressão

Coeficiente	Distribuição			Parâmetro
	GIW	Weibull	Exponencial	
<b>Mu</b>				
Intercepto	3.52707	4.43560	4.64403	Estimativa
	14.62542	0.95661	1.00430	EP
	0.241	4.637	4.624	$t$
Idade	0.810	1.25e-05	1.3e-05	$Pr(>  t )$
	-0.01997	-0.02162	-0.02336	Estimativa
	0.01275	0.01325	0.01427	EP
Estágio	-1.566	-1.632	-1.637	$t$
	0.121	0.106302	0.105207	$Pr(>  t )$
	-0.64711	-0.46781	-0.49445	Estimativa
	0.13816	0.12825	0.13529	EP
	-4.684	-3.648	-3.655	$t$
	1.06e-05	0.000453	0.000439	$Pr(>  t )$
<b>Sigma</b>				
Intercepto	-0.3725	0.09302	-	Estimativa
	0.1026	0.12088	-	EP
	-3.631	0.77	-	$t$
	0.000481	0.444	-	$Pr(>  t )$
<b>Nu</b>				
Intercepto	0.2091	-	-	Estimativa
	10.0452	-	-	EP
	0.021	-	-	$t$
	0.983	-	-	$Pr(>  t )$

Tendo em vista o p-valor, nota-se a insignificância estatística da covariável *idade* em todas as distribuições, fator que foi também observado na análise exploratória realizada, apresentando baixa correlação com o tempo para óbito dos indivíduos. De outro ângulo, *estagio* apresenta uma grande significância estatística, em especial na regressão realizada com a distribuição GIW.

Pode-se também observar o ajuste dos resíduos. Abaixo apresenta-se os resultados para a distribuição de interesse deste projeto.



**Fig. 16:** Análise de Resíduos da Regressão com GIW

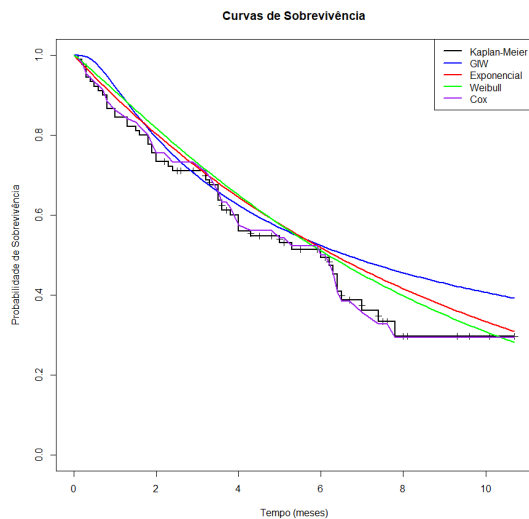
Vê-se que os resíduos têm uma densidade muito parecida com a da distribuição normal, o que também é evidenciado pelo Q-Q Plot, com os valores repousando próximos à linha. Isso indica que as suposições iniciais da regressão são cumpridas com o ajuste pela GIW com algumas variações.

Seguidamente, realizou-se a regressão com o modelo semi-paramétrico de Cox. Abaixo está um sumário dos resultados.

**TABLE IV: Resultados da Regressão**

Covariável	Coefficiente	exp(Coef)	EP	z	p-valor
idade	0.02279	1.02305	0.01439	1.583	0.113329
estagio	0.50143	1.65108	0.13948	3.595	0.000325

Novamente, a covariável *idade* é tida como insignificante para o modelo. Com isso, pode-se analisar as respectivas curvas de sobrevivência de cada modelo com base nos coeficientes estimados e compará-las com a curva de Kaplan-Meier:



**Fig. 17: Curvas de Sobrevivência com os Parâmetros das Regressões**

Com base na figura 17, é possível perceber que a regressão semi-paramétrica de Cox é a que melhor se ajusta aos dados observados, tal fator pode estar relacionado à sua flexibilidade em lidar com dados reais, algo que não ocorre com modelos paramétricos, cuja forma é bem definida. Também é possível observar que o modelo de regressão utilizando a distribuição Weibull Inversa Generalizada é o que obtém o pior desempenho, o que pode, em parte, estar relacionado à sua aplicabilidade de caráter industrial e epidemiológica, a qual não condiz com o contexto dos dados aplicados.

Por fim, tem-se abaixo uma comparação dos ajustes dos modelos a partir das métricas AIC, BIC e Global Deviance.

**TABLE V: Comparação de Modelos**

	EXP	WEI	GIW	Cox
AIC	291.91	293.366	298.7878	381.9035
BIC	299.4094	303.3653	311.2868	385.7275
Global Deviance	285.91	285.366	288.7878	-

Conforme esperado, o modelo exponencial é o que apresenta melhor ajuste, uma vez que é o mais simples. Já o modelo aqui estudado novamente tem um dos piores desempenhos, o que é compreensível dado suas características exploradas nas etapas anteriores e sua aplicabilidade.

## REFERENCES

- [1] W. Weibull, "A statistical distribution function of wide applicability," *Journal of Applied Mechanics*, vol. 18, pp. 293–297, 1951.
- [2] A. Drapella, "Complementary weibull distribution: unknown or just forgotten," *Quality and Reliability Engineering International*, vol. 9, pp. 383–385, 1993.
- [3] A. S. Hassan and S. G. Nassr, "The inverse weibull generator of distributions: Properties and applications," *Journal of Data Science*, vol. 16, no. 4, 2018.
- [4] S. Alkarni, A. Z. Afify, I. Elbatal, and M. Elgarhy, "The extended inverse weibull distribution: Properties and applications," *Complexity*, vol. 2020, pp. 1–11, 2020.
- [5] R. S. Felipe, M. M. Edwin, and M. C. Gauss, "The generalized inverse weibull distribution," *Statistical Papers*, vol. 52, no. 3, pp. 591–619, 2009.
- [6] F. R. S. d. GUSMÃO *et al.*, "Uma abordagem bayesiana para distribuição weibull inversa generalizada," 2008.
- [7] F. Mukhamedov, K. Jain, N. Singla, and S. K. Sharma, "The generalized inverse generalized weibull distribution and its properties," *Journal of Probability*, 2014.
- [8] S. Tuli, S. Tuli, R. Tuli, and S. S. Gill, "Predicting the growth and trend of covid-19 pandemic using machine learning and cloud computing," *Internet of things*, vol. 11, p. 100222, 2020.
- [9] T. N. Sindhu, A. Shafiq, and Q. M. Al-Mdallal, "Exponentiated transformation of gumbel type-ii distribution for modeling covid-19 data," *Alexandria Engineering Journal*, vol. 60, no. 1, pp. 671–689, 2021.
- [10] —, "On the analysis of number of deaths due to covid- 19 outbreak data using a new class of distributions," *Results in Physics*, vol. 21, p. 103747, 2021.
- [11] P. S. N. BARROS *et al.*, "Classes de distribuições weibull generalizada: teorias e aplicações," 2015.
- [12] D. L. ROGES *et al.*, "A distribuição beta weibull inversa generalizada: desenvolvimentos e aplicações," 2011.
- [13] V. S. Melo, "Modelos envolvendo duas distribuições weibull inversa e apresentação da weibull inversa generalizada," 2014.
- [14] S. J. Almalki and S. Nadarajah, "Modifications of the weibull distribution: A review," *Reliability Engineering & System Safety*, vol. 124, pp. 32–55, 2014.
- [15] E. A. Colosimo and S. R. Giolo, *Análise de sobrevivência aplicada*. Editora Blucher, 2021.
- [16] M. A. M. Salih and Z. S. Noori, "Using simulation to estimate reliability function for generalized inverse weibull distribution," *Journal of Physics: Conference Series*, vol. 2322, no. 1, p. 012013, aug 2022. [Online]. Available: <https://dx.doi.org/10.1088/1742-6596/2322/1/012013>



## APPENDIX

## A. Códigos da Etapa 3

```

1 # Load necessary libraries
2 set.seed(123)
3 library(gamlss)
4 library(RelDists)
5
6 # Function to generate data, calculate statistics, and plot histogram
7 generate_analyze_and_plot <- function(alpha, beta, gamma, sample_size) {
8   # Generate random numbers from the GIW distribution
9   simulated_data <- rGIW(n = sample_size, mu = alpha, sigma = beta, nu = gamma)
10
11   # Calculate statistics
12   quantiles <- quantile(simulated_data, probs = c(0.25, 0.5, 0.75))
13   estimated_q025 <- quantile(simulated_data, probs = 0.25)
14   estimated_q075 <- quantile(simulated_data, probs = 0.75)
15   estimated_mean <- mean(simulated_data)
16   estimated_variance <- var(simulated_data)
17   median_value <- median(simulated_data)
18
19   # real_quantiles
20   true_q025 = alpha*(-gamma/log(0.25))^(1/beta)
21   true_median = alpha*(-gamma/log(0.5))^(1/beta) # Mediana
22   true_q075 = alpha*(-gamma/log(0.75))^(1/beta)
23
24   # Real values for mean and variance
25   true_mean <- alpha * gamma(1 - 1/beta) * (gamma^(1/beta))
26   true_variance <- (gamma^(2/beta)) * (alpha^2) * gamma(1 - 2/beta) - (true_mean^2)
27
28   # Fit GIW distribution to the data to estimate parameters
29   mod <- gamlss(simulated_data ~ 1, family = GIW(mu.link = "log", sigma.link = "log", nu.link =
30     "log"))
31   estimated_mu <- exp(coef(mod, what = "mu"))
32   estimated_sigma <- exp(coef(mod, what = "sigma"))
33   estimated_nu <- exp(coef(mod, what = "nu"))
34
35   # Calculate bias and MSE
36   bias_mean <- abs(estimated_mean - true_mean)
37   bias_variance <- abs(estimated_variance - true_variance)
38   bias_q025 <- abs(estimated_q025 - true_q025)
39   bias_median <- abs(median_value - true_median)
40   bias_q075 <- abs(estimated_q075 - true_q075)
41   mse_mean <- (estimated_mean - true_mean)^2 + true_variance/n
42   mse_variance <- (estimated_variance - true_variance)^2 + (true_variance^4)*(2*(n-1))/n^2
43   bias_alpha <- abs(estimated_mu - alpha)
44   bias_beta <- abs(estimated_sigma - beta)
45   bias_gamma <- abs(estimated_nu - gamma)
46
47   # Plot histogram with relative frequencies
48   hist(simulated_data, breaks = 30, main = paste("Histogram of Simulated GIW Data\nAlpha:",
49     alpha, "Beta:", beta, "Gamma:", gamma, "Sample Size:", sample_size), xlab = "Value", freq
50     = FALSE)
51
52   # Fit a curve with true parameters
53   curve(dGIW(x, alpha, beta, gamma), add = TRUE, col = "red")
54
55   # Fit a curve with estimated parameters
56   curve(dGIW(x, estimated_mu, estimated_sigma, estimated_nu), add = TRUE, col = "blue", lwd =
57     2, lty = 2)
58
59   # Return the calculated statistics and estimated parameters
60   return(list(quantiles = quantiles, mean = estimated_mean, median = median_value, variance =
61     estimated_variance,
62     bias_mean = bias_mean, bias_variance = bias_variance, bias_q025 = bias_q025, bias
63     _median = bias_median,
64     bias_q075 = bias_q075, mse_mean = mse_mean, mse_variance = mse_variance,
65     estimated_mu = estimated_mu,

```

```

59         estimated_sigma = estimated_sigma, estimated_nu = estimated_nu, bias_alpha = bias
60         _alpha, bias_beta = bias_beta,
        bias_gama = bias_gama))
61     }
62
63     # Define values for alpha, beta, gamma, and sample_size
64     alpha <- c(2, 4, 7)
65     beta <- c(5, 10, 15)
66     gamma <- c(1, 2, 4)
67     sample_size <- c(100, 500, 2000)
68
69     # Iterate over all combinations of parameters
70     for (a in alpha) {
71         for (b in beta) {
72             for (g in gamma) {
73                 for (s in sample_size) {
74                     # Generate, analyze, and plot data
75                     stats <- generate_analyze_and_plot(a, b, g, s)
76
77                     # Print results
78                     cat("Alpha:", a, "Beta:", b, "Gamma:", g, "Sample Size:", s, "\n")
79                     cat("Quantiles (0.25, 0.5, 0.75):", stats$quantiles, "\n")
80                     cat("Mean:", stats$mean, "\n")
81                     cat("Median:", stats$median, "\n")
82                     cat("Variance:", stats$variance, "\n")
83                     cat("Bias (Mean):", stats$bias_mean, "\n")
84                     cat("Bias (Variance):", stats$bias_variance, "\n")
85                     cat("Bias (q025):", stats$bias_q025, "\n")
86                     cat("Bias (Median):", stats$bias_median, "\n")
87                     cat("Bias (q075):", stats$bias_q075, "\n")
88                     cat("MSE (Mean):", stats$mse_mean, "\n")
89                     cat("MSE (Variance):", stats$mse_variance, "\n")
90                     cat("Estimated Mu:", stats$estimated_mu, "\n")
91                     cat("Estimated Sigma:", stats$estimated_sigma, "\n")
92                     cat("Estimated Nu:", stats$estimated_nu, "\n")
93                     cat("Bias (Mu):", stats$bias_alpha, "\n")
94                     cat("Bias (Sigma):", stats$bias_beta, "\n")
95                     cat("Bias (Nu):", stats$bias_gama, "\n\n")
96                 }
97             }
98         }
99     }

```

**Listing 1:** Código para simulação da GIW e estimação de parâmetros