

Comparação e Desempenhos dos Testes de Normalidade via Simulação de Monte Carlo em Linguagem R

Mário Diego Rocha Valente (Graduando em Sistemas de Informação da UFPA)

Breno Cauã Rodrigues da Silva (Graduando em Estatística da UFPA)

Dennison Célio de Oliveira Carvalho (Prof. Dr., UFPA)

Heliton Ribeiro Tavares (Prof. Dr., UFPA)

Email:diego.valente@icen.ufpa.br, breno.silva@icen.ufpa.br, dennison@ufpa.br, heliton@ufpa.br



1. INTRODUÇÃO

De modo geral, muitos métodos estatísticos requerem que os dados sigam uma distribuição normal e que as observações sejam independentes. Verificar a proximidade da distribuição dos dados com a normalidade torna-se, portanto, uma tarefa fundamental para garantir a validade desses métodos. Mas como podemos assegurar que a distribuição dos dados se aproxima de uma distribuição normal? Duas métricas primordiais para avaliar o desempenho dos testes de normalidade são as **Taxas de Erro Tipo I**, calculadas sob a hipótese nula (H_0), e o **Poder do Teste**, que é mensurado sob a hipótese alternativa (H_1). A análise dessas métricas permite identificar a capacidade de cada teste em evitar erros de falso positivo e, ao mesmo tempo, sua eficácia em detectar desvios de normalidade quando eles de fato existem (CARDOSO e FERREIRA, 2010).

2. MATERIAIS E MÉTODOS

2.1 Software

Para conduzir as análises e estimativas neste estudo, foi utilizada a linguagem de programação $R_{4.5.1}$, empregando a IDE **RStudio**, versão 1.12. Os seguintes pacotes foram utilizados nas diversas etapas do processo de simulação dos testes de normalidade: **DistributionTest**, **irr**, **moments**, **nortest**, **DescTools**.

2.2 Simulação de Monte Carlo

Neste estudo, a eficácia dos testes de normalidade foi avaliada por meio da simulação de amostras tanto de distribuições normais quanto não normais, abrangendo casos simétricos e assimétricos, incluindo as distribuições Normal, Uniforme, Beta, t-Student, Cauchy e Exponencial. Foram considerados seis tamanhos amostrais distintos para cada distribuição: 30, 50, 100, 300, 500 e 1000. Em cada caso, foram realizadas 2.000 replicações, durante as quais os testes de normalidade foram registrados para análise posterior. Utilizou-se o teste Kappa de Fleiss, implementado pela função `Kapam.fleiss()` do pacote `irr`, com o objetivo de avaliar o nível de concordância entre os testes de normalidade.

3. RESULTADOS E DISCUSSÃO

O script em R foi armazenado no https://github.com/MarioDhiego/Teste_Normalidade.

Tabela 1 Teste kappa-Fleiss de Concordância para os testes de normalidade.

Distribuição de Probabilidade	Teste de Kappa-Fleiss
Normal(0, 1)	0,93
Beta(2, 5)	0,89
Cauchy(0, 1)	0,81

Tabela 2 Taxa Média de Acertos para a Sensibilidade dos Testes de Normalidade.

Distribuição	Amostra	AD	CVM	DA	JB	KS	Lill	SW	Z_A	Z_C	Z_K
Beta (2,5)	30	1.0	0.82	0.81	0.87	1.0	0.85	0.75	0.73	0.76	0.81
	50	1.0	0.74	0.73	0.84	0.99	0.79	0.62	0.55	0.63	0.68
	100	0.99	0.51	0.47	0.70	0.98	0.62	0.27	0.17	0.27	0.28
	300	0.99	0.49	0.46	0.69	0.98	0.61	0.27	0.18	0.26	0.29
	500	0.28	0.003	0.001	0.002	0.49	0.02	0.01	0.01	0.01	0.01
	1000	0.004	0.10	0.010	0.10	0.10	0.10	0.10	0.10	0.10	0.10
Normal (0,1)	30	1.0	0.95	0.94	0.93	1.0	0.95	0.94	0.94	0.94	0.95
	50	1.0	0.94	0.94	0.93	1.0	0.94	0.94	0.94	0.94	0.94
	100	1.0	0.95	0.95	0.93	1.0	0.95	0.95	0.94	0.94	0.94
	300	1.0	0.95	0.95	0.94	1.0	0.94	0.95	0.95	0.95	0.95
	500	1.0	0.95	0.95	0.95	1.0	0.95	0.95	0.95	0.95	0.95
	1000	1.0	0.96	0.96	0.96	1.0	0.96	0.95	0.96	0.96	0.96
Cauchy (0,1)	30	0.30	0.05	0.16	0.03	0.32	0.08	0.06	0.07	0.08	0.07
	50	0.19	0.02	0.12	0.01	0.19	0.03	0.02	0.02	0.03	0.03
	100	0.08	0.08	0.08	0.05	0.08	0.01	0.01	0.01	0.01	0.02
	300	0.04	0.01	0.03	0.01	0.05	0.01	0.01	0.01	0.01	0.01
	500	0.03	0.01	0.02	0.01	0.03	0.01	0.01	0.01	0.01	0.01
	1000	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

4. SIMULAÇÃO em LINGUAGEM R

4.1 Script



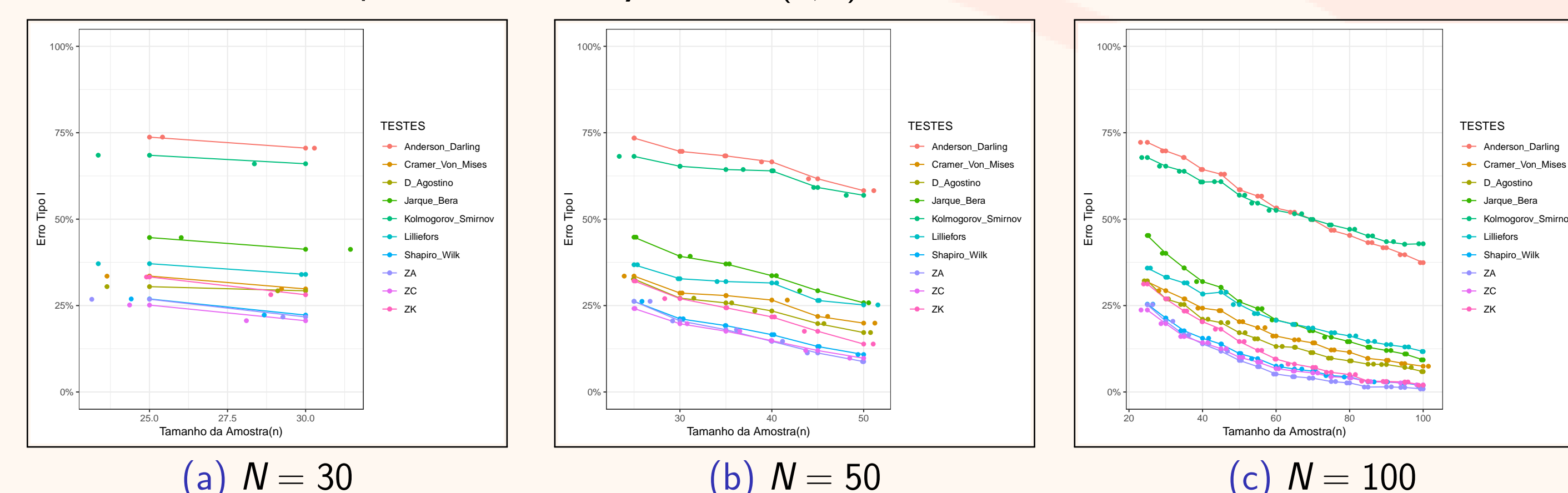
(a) Tamanho de Amostra

(b) Erro Tipo I (P-valor)

(c) Poder do Teste

Figure 1: Simulações em linguagem R: (a) Tamanho de amostra, (b) Erro Tipo I e (c) Terceira simulação.

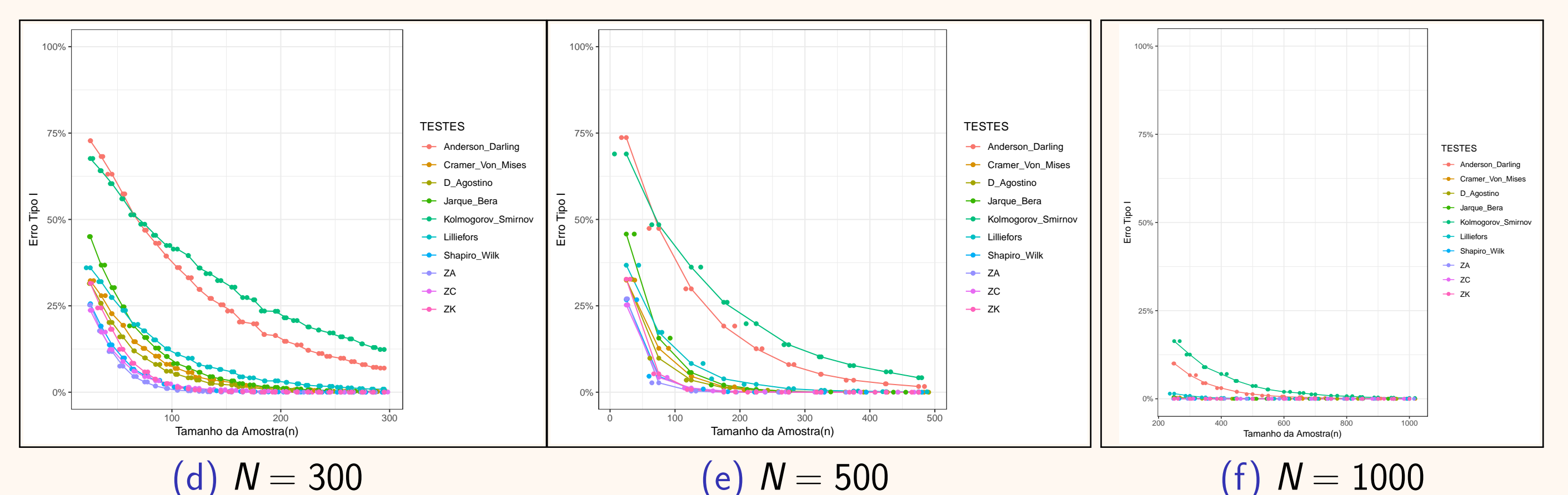
Figure 2: Comparação do Erro Tipo I dos testes AD, CM, DG, LL, JB, KS, ZA, ZC e ZK em função do tamanho amostral para a Distribuição Beta(2, 5).



(a) N = 30

(b) N = 50

(c) N = 100

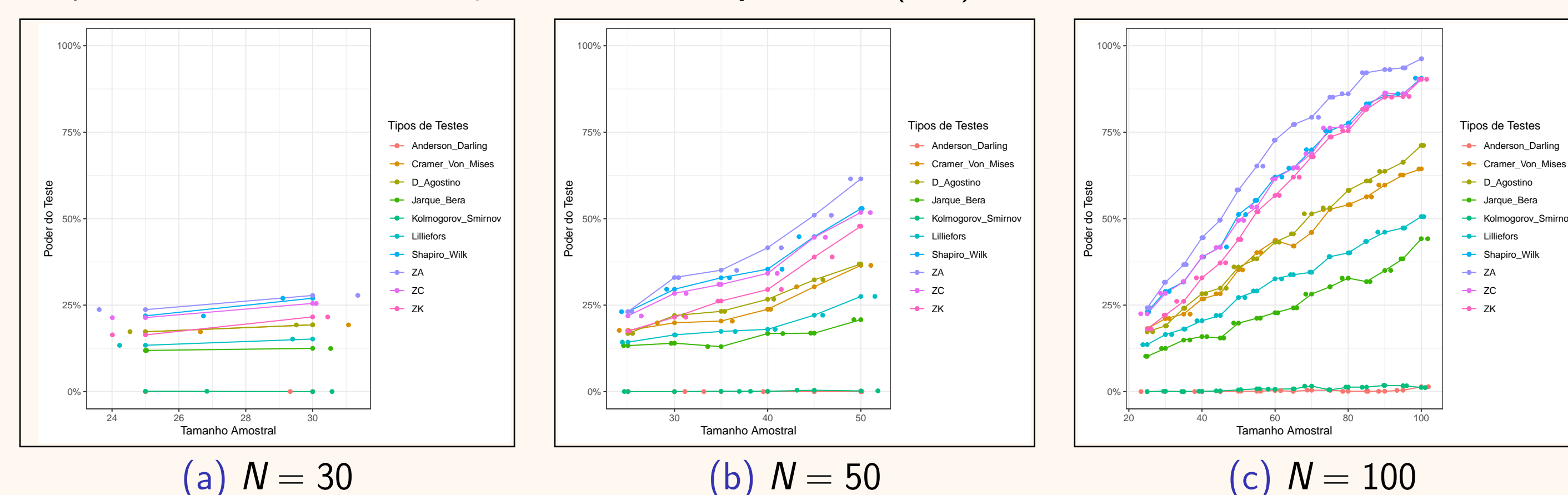


(d) N = 300

(e) N = 500

(f) N = 1000

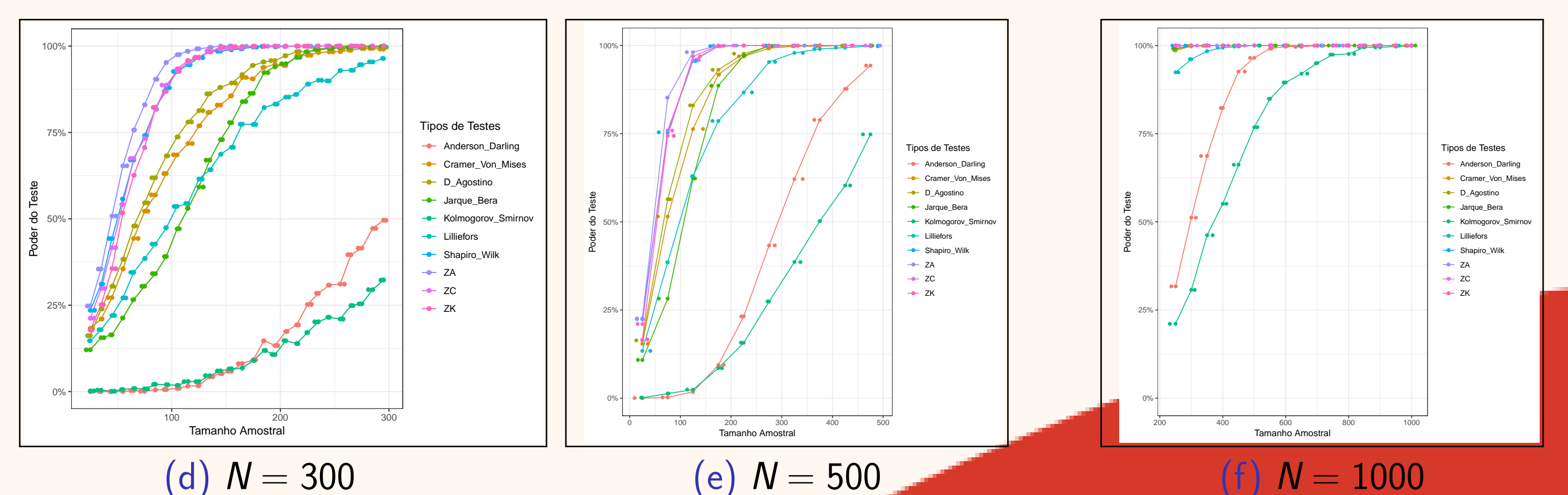
Figure 3: Comparação do Poder do Teste dos testes AD, CM, DG, LL, JB, KS, ZA, ZC e ZK em função do tamanho amostral para a Distribuição Beta(2, 5).



(a) N = 30

(b) N = 50

(c) N = 100



(d) N = 300

(e) N = 500

(f) N = 1000

5. CONSIDERAÇÕES FINAIS

Para amostras maiores que 500, o teste de Zhang (2002) apresentou o melhor desempenho. Em amostras menores, o teste de Shapiro-Wilk obteve os melhores resultados. O teste de Kappa-Fleiss foi aplicado, e na maioria das distribuições, os testes tiveram forte concordância. Por outro lado, os testes de KS e AD apresentaram baixa potência, mesmo em amostras grandes.

