

Este documento consiste em uma série de ensaios que descrevem a pesquisa conduzida pelos estatísticos para desenvolver os métodos e verificações de dados usados no Assistente no Minitab Statistical Software.

Teste t para 1 Amostra

Visão geral

O teste t para 1 amostra é usado para estimar a média do processo e para compará-la a um valor alvo. Esse teste é considerado um procedimento robusto por ser extremamente insensível à suposição de normalidade quando a amostra é moderadamente grande. Segundo a maioria dos compêndios estatísticos, o teste t para 1 amostra e o intervalo de confiança para a média são apropriados para amostras de tamanho 30 ou maiores.

Neste documento, descrevemos as simulações conduzidas para avaliar a regra geral de mínimo de 30 unidades de amostra. Nossas simulações focaram no impacto da não normalidade sobre o teste t para 1 amostra. Além disso, queríamos avaliar o impacto dos dados atípicos sobre os resultados do teste.

Com base em nossa pesquisa, o Assistente realiza automaticamente as seguintes verificações nos dados e exibe os resultados no Report Card:

- Dados atípicos
- Normalidade (A amostra é suficientemente grande para que a normalidade não seja um problema?)
- Tamanho amostral

Para informações gerais sobre a metodologia do teste t para 1 amostra, consulte Arnold (1990), Casella e Berger (1990), Moore e McCabe (1993) e Srivastava (1958).

Nota Os resultados neste documento se aplicam também ao teste t pareado no Assistente, já que o teste t pareado aplica o mesmo método do teste t para 1 amostra às amostras de diferenças pareadas.

Verificações de dados

Dados atípicos

Dados atípicos são valores de dados extremamente grandes ou pequenos, também conhecidos como outliers. Dados atípicos podem ter forte influência sobre os resultados da análise. Quando a amostra é pequena, eles podem afetar as chances de se identificar resultados significativos estatisticamente. Dados atípicos podem indicar problemas com coleta de dados ou um comportamento atípico do processo que você está estudando. Esses pontos de dados geralmente devem ser investigados e corrigidos quando possível.

Objetivo

Queríamos desenvolver um método para verificar valores de dados que fossem muito grandes ou muito pequenos na amostra geral, e que podem afetar o resultado da análise.



Método

Desenvolvemos um método de verificação de dados atípicos que se baseia no método descrito por Hoaglin, Iglewicz e Tukey (1986) para identificar outliers nos boxplots.

Resultados

O Assistente identifica um ponto de dados como atípico quando sua amplitude interquartílica ultrapassa em 1,5 vez o quartil inferior ou superior da distribuição. Os quartis superior e inferior correspondem aos percentis 25° e 75° dos dados. A amplitude interquartílica é a diferença entre os dois quartis. Esse método funciona bem mesmo quando existem vários outliers porque possibilita detectar cada outlier específico.

Ao verificar dados atípicos, o Assistente de Report Card de teste t para 1 amostra exibe os seguintes indicadores de status:

Status	Condição
	Não há pontos de dados atípicos.
	Existe pelo menos um ponto de dados atípico que pode afetar os resultados do teste.

Normalidade

O teste t para 1 amostra parte da suposição de que a população é normalmente distribuída. Felizmente, mesmo se os dados não estiverem distribuídos normalmente, esse método funcionará bem quando o tamanho de amostra for suficientemente grande.

Objetivo

Queríamos determinar o efeito da não normalidade sobre os erros tipo I e tipo II do teste com o intuito de fornecer diretrizes sobre o tamanho e normalidade da amostra.

Método

Conduzimos simulações para determinar o tamanho de amostra que poderia ser ignorado pela suposição de normalidade durante a realização de um teste t para 1 amostra ou durante o cálculo de um intervalo de confiança para 1 amostra voltado à média de uma população.

Planejamos o primeiro estudo para avaliar o efeito da não normalidade sobre a taxa de erro tipo I do teste. Especificamente, queríamos deduzir o tamanho mínimo de amostra exigido no teste de modo que fosse insensível à distribuição da população. Realizamos o teste t para 1 amostra em amostras pequenas, moderadas e grandes geradas a partir de populações normais e não normais. As populações não normais incluíam populações leve ou severamente assimétricas, populações unilaterais pouco ou extremamente simétricas e populações normais contaminadas. A população normal serviu como população de controle para fins de comparação. Para cada caso, calculamos os níveis de significância simulados e os comparamos com o nível de significância nominal, desejado, de 0,05. Se o teste for bem realizado, os níveis de significância simulados serão próximos a 0,05. Examinamos os níveis de significância simulados em todas as diferentes condições para avaliar o tamanho mínimo de amostra na qual eles permaneceriam próximos ao nível alvo, independentemente da distribuição. Para mais detalhes, consulte o apêndice A.



No segundo estudo, examinamos o efeito da não normalidade sobre o erro tipo II do teste. O experimento de simulação é idêntico ao do primeiro estudo. No entanto, comparamos os níveis de poder simulados sob diferentes condições com os níveis de poder alvo calculados por meio da função de poder teórica do teste t para 1 amostra. Para obter mais detalhes, consulte o apêndice B.

Resultados

O efeito da não normalidade sobre ambas as taxas de erro de tipo I e II do teste é mínimo para tamanhos de amostra pequenos, como 20. Entretanto, quando a população parental da amostra é extremamente assimétrica, amostras maiores podem ser solicitadas.

Recomendamos um tamanho de amostra de aproximadamente 40 para esses casos. Para obter mais detalhes, consulte o apêndice A e o apêndice B.

Como o teste mostrou bom desempenho com amostras relativamente pequenas, o Assistente não testou a normalidade os dados. Em vez disso, ele verifica o tamanho da amostra e exibe os seguintes indicadores de status no Report Card:

Status	Condição
	O tamanho da amostra é de pelo menos 20; normalidade não é um problema.
	O tamanho da amostra é menor do que 20; normalidade pode ser um problema.

Tamanho amostral

Geralmente, é realizado um teste de hipóteses para reunir evidências que permitam rejeitar a hipótese nula de “nenhuma diferença”. Se as amostras forem muito pequenas, é possível que o poder de teste não seja adequado para detectar uma diferença real existente entre as médias, o que resulta em um erro tipo II. Desse modo, é crucial garantir que o tamanho de amostra seja suficientemente grande para detectar diferenças realmente importantes, com alta probabilidade.

Objetivo

Se os dados não fornecerem evidência suficiente contra a hipótese nula, queremos determinar se o tamanho de amostra é suficientemente grande para que o teste detecte diferenças reais de interesse, com alta probabilidade. Embora o objetivo do planejamento de tamanho amostral consista em assegurar que o tamanho de amostra seja suficientemente grande para detectar diferenças importantes com alta probabilidade, ele não deveria ser grande a ponto de fazer com que diferenças menos significativas se tornem estatisticamente significativas com alta probabilidade.

Método






A análise de tamanho amostral e poder é baseada na função de poder teórica do teste específico usado para conduzir a análise estatística. Conforme analisado anteriormente, a função de poder do teste t para 1 amostra é insensível à suposição normal quando o tamanho de amostra tiver um valor de, no mínimo, 20. A função de poder depende do tamanho da amostra, da diferença entre a média alvo e a média da população e da variância da população. Para obter mais detalhes, consulte o apêndice B.

Resultados

Quando os dados não fornecem evidência suficiente contra a hipótese nula, o Assistente calcula as diferenças reais que podem ser detectadas com níveis de probabilidades de 80% e 90% para os tamanhos amostrais fornecidos. Além disso, se o usuário fornece uma diferença de interesse real em particular, o Assistente calcula um tamanho de amostra que gere uma chance de 80% e de 90% de detecção da diferença.

Não existe um resultado geral a relatar porque os resultados dependem da amostra específica ao usuário. Consulte o apêndice B se quiser mais informações sobre poder em testes t para 1 amostra.

Ao verificar o poder e o tamanho amostral, o Assistente de Report Card do teste t para 1 amostra exibe os seguintes indicadores de status:

Status	Condição
	<p>O teste encontra uma diferença entre a média e o alvo, desse modo, o poder não é um problema.</p> <p>OU</p> <p>O poder é suficiente. O teste não encontrou nenhuma diferença entre a média e o alvo, mas a amostra é suficientemente grande para fornecer um mínimo de 90% de chance de detectar a diferença fornecida.</p>
	<p>O poder pode ser suficiente. O teste não encontrou nenhuma diferença entre a média e o alvo, mas a amostra é suficientemente grande para fornecer 80% a 90% de chance de detectar a diferença fornecida. É relatado o tamanho de amostra exigido para atingir um poder de 90%.</p>
	<p>Talvez o poder não seja suficiente. O teste não encontrou nenhuma diferença entre a média e o alvo, e a amostra é suficientemente grande para fornecer entre 60% e 80% de chance de detectar a diferença fornecida. São relatados os tamanhos amostrais necessários para alcançar 80% e 90% de poder.</p>
	<p>O poder não é suficiente. O teste não encontrou nenhuma diferença entre a média e o alvo, e a amostra não é suficientemente grande para fornecer um mínimo de 60% de chance de detectar a diferença fornecida. São relatados os tamanhos amostrais necessários para alcançar 80% e 90% de poder.</p>
	<p>O teste não encontrou nenhuma diferença entre a média e o alvo. Você não especificou nenhuma diferença real a ser detectada entre a média e o alvo; desse modo, o relatório indica as diferenças que poderiam ser detectadas com 80% e 90% de chance, com base nos tamanhos amostrais, desvios padrão e alfa.</p>

Referências

- Arnold, S.F. (1990). *Mathematical statistics*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Casella, G., e Berger, R. L. (1990). *Statistical inference*. Pacific Grove, CA: Wadsworth, Inc.
- Hoaglin, D. C., Iglewicz, B. e Tukey, J. W. (1986). Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association*, 81, 991-999.
- Moore, D.S. e McCabe, G.P. (1993). *Introduction to the practice of statistics*, 2nd ed. New York, NY: W. H. Freeman and Company.
- Neyman, J., Iwazkiewicz, K. e Kolodziejczyk, S. (1935). Statistical problems in agricultural experimentation, *Journal of the Royal Statistical Society, Series B*, 2, 107-180.
- Pearson, E.S., e Hartley, H.O. (Eds.). (1954). *Biometrika tables for statisticians*, Vol. I. London: Cambridge University Press.
- Srivastava, A. B. L. (1958). Effect of non-normality on the power function of t-test, *Biometrika*, 45, 421-429.

Apêndice A: Impacto da não normalidade sobre o nível de significância (validade do teste)

Sob a suposição normal, o teste t para 1 amostra é um teste de tamanho α não viciado UMP (*uniformly most powerful*). Ou seja, o teste é tão poderoso quanto ou ainda mais poderoso que qualquer outro teste de tamanho α não viciado sobre a média. No entanto, quando a população parental da amostra não estiver distribuída normalmente, as propriedades de otimalidade serão confirmadas se o tamanho da amostra for grande o suficiente. Em outras palavras, nas amostras suficientemente grandes, o nível real de significância do teste t para 1 amostra praticamente se iguala ao nível alvo de dados normais e não-normais, e a função poder do teste também é insensível à suposição normal (Srivastava, 1958).

Queríamos determinar o quão suficientemente grande precisa ser uma amostra para que o teste t seja insensível à suposição normal. Muitos compêndios recomendam que, se o tamanho amostral $n \geq 30$, a suposição normal poderá ser ignorada para a maioria das finalidades práticas (Arnold, 1990; Casella e Berger, 1990; e Moore e McCabe, 1993). A finalidade da investigação descrita nesses apêndices é conduzir estudos de simulação para avaliar a regra geral, examinando o impacto das diferentes distribuições não normais sobre o teste t para 1 amostra.

Estudo de simulação A

Queríamos examinar o impacto da não normalidade sobre a taxa de erro de tipo I do teste para avaliar o tamanho de amostra mínimo no qual ela seja estável e permaneça próxima à taxa de erro alvo, independentemente da distribuição.

Para conseguir isso, foram realizados testes t bilaterais com $\alpha = 0,05$ usando amostras aleatórias de vários tamanhos ($n = 10, 15, 20, 25, 30, 35, 40, 50, 60, 80, 100$) geradas a partir de distribuições com diferentes propriedades. Tais distribuições incluem

- A distribuição normal padrão ($N(0, 1)$)
- Distribuições simétricas com caudas pesadas, como a distribuição t com 5 e 10 graus de liberdade ($t(5), t(10)$)
- A distribuição de Laplace com locação 0 e escala 1 (Lpl)
- Distribuições assimétricas de caudas pesadas representadas pela distribuição exponencial com escala 1 (Exp), as distribuições Qui-quadrado com 3, 5 e 10 graus de liberdade ($Chi(3), Chi(5), Chi(10)$)

- Distribuições simétricas e de caudas mais leves, como a distribuição uniforme (U(0, 1)), a distribuição Beta com os dois parâmetros definidos como 3 (B(3, 3))
- Uma distribuição assimétrica (inclinada) para a esquerda e com cauda pesada (B(8, 1))

Além disso, para avaliar o efeito direto dos outliers, geramos amostras a partir de distribuições normais contaminadas definidas como:

$$CN(p, \sigma) = pN(0, 1) + (1 - p)N(0, \sigma)$$

onde p é definido como o parâmetro de mistura, e $1 - p$ é a proporção de contaminação ou proporção de outliers. Selecionamos duas populações normais contaminadas para o estudo: CN(0,9, 3) (10% dos membros da população são outliers) e CN(0,8, 3) (20% dos membros da população são outliers). Essas duas distribuições são simétricas e têm caudas (extremidades) longas devido aos outliers.

Para cada tamanho amostral, foram obtidas réplicas de 10.000 amostras de cada população, e foi realizado um teste t para 1 amostra com hipótese nula $\mu = \mu_o$ e hipótese alternativa $\mu \neq \mu_o$ para cada uma das 10.000 amostras. Para cada teste, definimos a média hipotética μ_o para a média verdadeira da população parental da amostra. Como resultado, para determinado tamanho amostral, a fração de réplicas de 10.000 amostras que geram uma rejeição da hipótese nula representa a taxa de erro de tipo I ou o nível de significância do teste. Como o nível de significância alvo é 5%, o erro de simulação é de aproximadamente 0,2%.

Os resultados da simulação são mostrados nas tabelas 1 e 2 e exibidos graficamente nas figuras 1 e 2.

Resultados e sumário

Os resultados (consulte a tabela 1 e a figura 1) mostram que mesmo nos tamanhos de amostra pequenos, quando as amostras são geradas a partir de populações simétricas os níveis de significância simulados do teste são próximos ao nível de significância alvo. Entretanto, os resultados do teste são levemente conservadores para as distribuições simétricas com cauda pesada quando se analisa amostras pequenas, incluindo as amostras pequenas geradas a partir de distribuições normais contaminadas. Percebe-se também que os outliers reduzem o nível de significância do teste quando as amostras são pequenas. Porém, tal efeito é revertido quando amostras pequenas são geradas de populações parentais de caudas mais leves (distribuições uniformes e Beta (3,3)). Os níveis de significância simulados são um pouco mais altos para esses casos.

Tabela 1 Níveis de significância simulados de teste t para 1 amostra bilateral em amostras geradas a partir de populações simétricas. O nível de significância alvo é $\alpha = 0,05$.

Dist.	N(0, 1)	t(5)	t(10)	Lpl	CN(0,9, 3)	CN(0,8, 3)	B(3, 3)	U(0, 1)
N	Caudas simétricas e pesadas						Caudas simétricas e mais leves	
10	0,050	0,046	0,048	0,044	0,043	0,039	0,057	0,057
15	0,051	0,050	0,049	0,049	0,043	0,043	0,053	0,054

Dist.	N(0, 1)	t(5)	t(10)	Lpl	CN(0,9, 3)	CN(0,8, 3)	B(3, 3)	U(0, 1)
N	Caudas simétricas e pesadas						Caudas simétricas e mais leves	
20	0,047	0,051	0,051	0,047	0,043	0,044	0,051	0,052
25	0,050	0,047	0,050	0,046	0,046	0,046	0,048	0,050
30	0,053	0,050	0,048	0,043	0,049	0,046	0,050	0,048
35	0,052	0,047	0,049	0,050	0,047	0,045	0,051	0,054
40	0,046	0,052	0,054	0,048	0,046	0,049	0,044	0,050
50	0,050	0,049	0,051	0,048	0,047	0,051	0,053	0,050
60	0,052	0,049	0,053	0,050	0,051	0,056	0,054	0,052
80	0,049	0,050	0,051	0,047	0,047	0,052	0,049	0,049
100	0,050	0,052	0,049	0,051	0,052	0,054	0,051	0,054

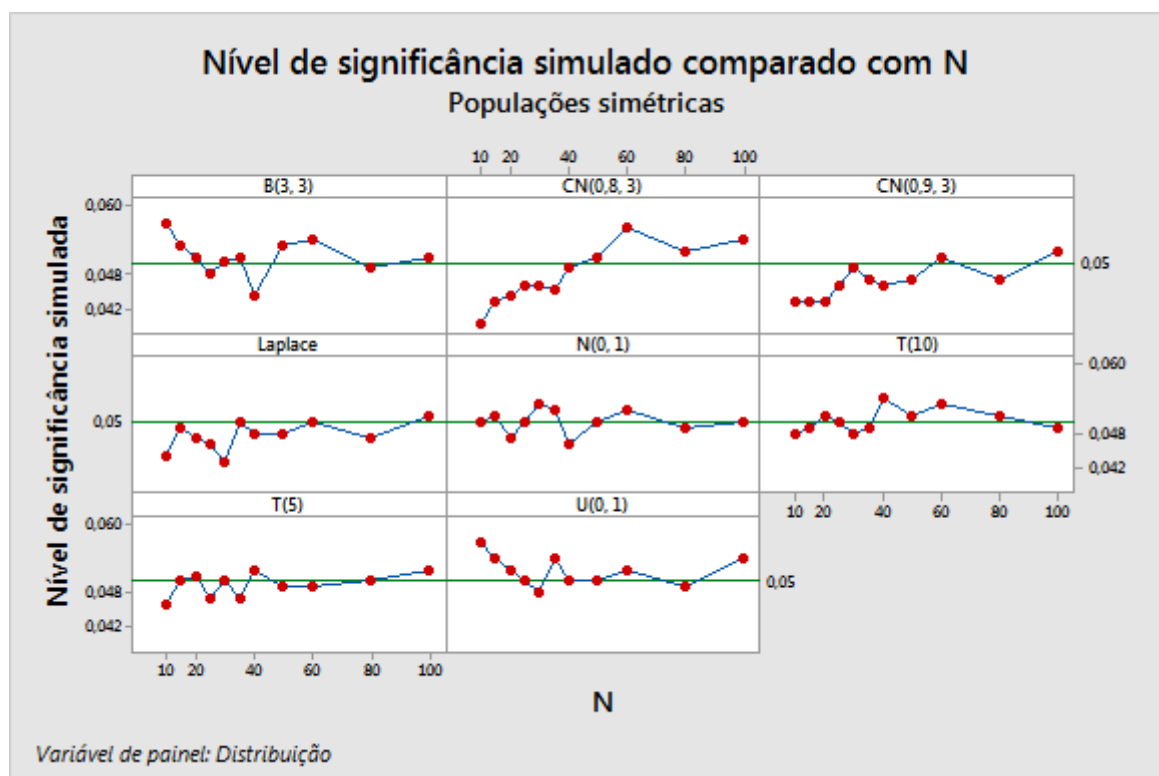


Figura 1 Gráfico de níveis de significância simulados de teste t para 1 amostra bilateral versus tamanho de amostras geradas a partir de populações simétricas. O nível de significância alvo é $\alpha = 0,05$.

Por outro lado, quando as amostras são geradas a partir de distribuições assimétricas, o desempenho do teste depende da gravidade da assimetria. Os resultados na tabela 2 e na figura 2 mostram que o teste t para 1 amostra é sensível à assimetria nas amostras

pequenas. No caso de populações severamente assimétricas (exponencial, Chi(3) e Beta(8, 1)), são necessárias amostras maiores para que os níveis de significância estejam próximos ao nível de significância alvo. No entanto, nas populações moderadamente assimétricas (Chi(5) e Chi(10)), um tamanho de amostra mínimo de 20 é suficiente para que os níveis de significância simulados se quedem próximos ao nível alvo. Com um tamanho de amostra de 20, o nível de significância simulado é aproximadamente 0,063 para a distribuição Qui-quadrado com 5 graus de liberdade e cerca de 0,056 para a distribuição Qui-quadrado com 10 graus de liberdade.

Tabela 2 Níveis de significância simulados de teste t para 1 amostra bilateral de amostras geradas a partir de populações assimétricas. O nível de significância alvo é $\alpha = 0,05$.

N	Exp	Chi(3)	B(8, 1)	Chi(5)	Chi(10)
	Assimetria da população				
	2,0	1,633	-1,423	1,265	0,894
	Níveis de significância simulados				
10	0,101	0,089	0,087	0,069	0,060
15	0,088	0,076	0,072	0,068	0,057
20	0,083	0,073	0,069	0,063	0,056
25	0,075	0,068	0,067	0,067	0,056
30	0,069	0,067	0,066	0,058	0,054
35	0,075	0,067	0,062	0,062	0,056
40	0,067	0,067	0,061	0,059	0,056
50	0,064	0,057	0,062	0,057	0,054
60	0,063	0,056	0,061	0,054	0,055
80	0,059	0,058	0,053	0,052	0,052
100	0,060	0,055	0,055	0,047	0,053

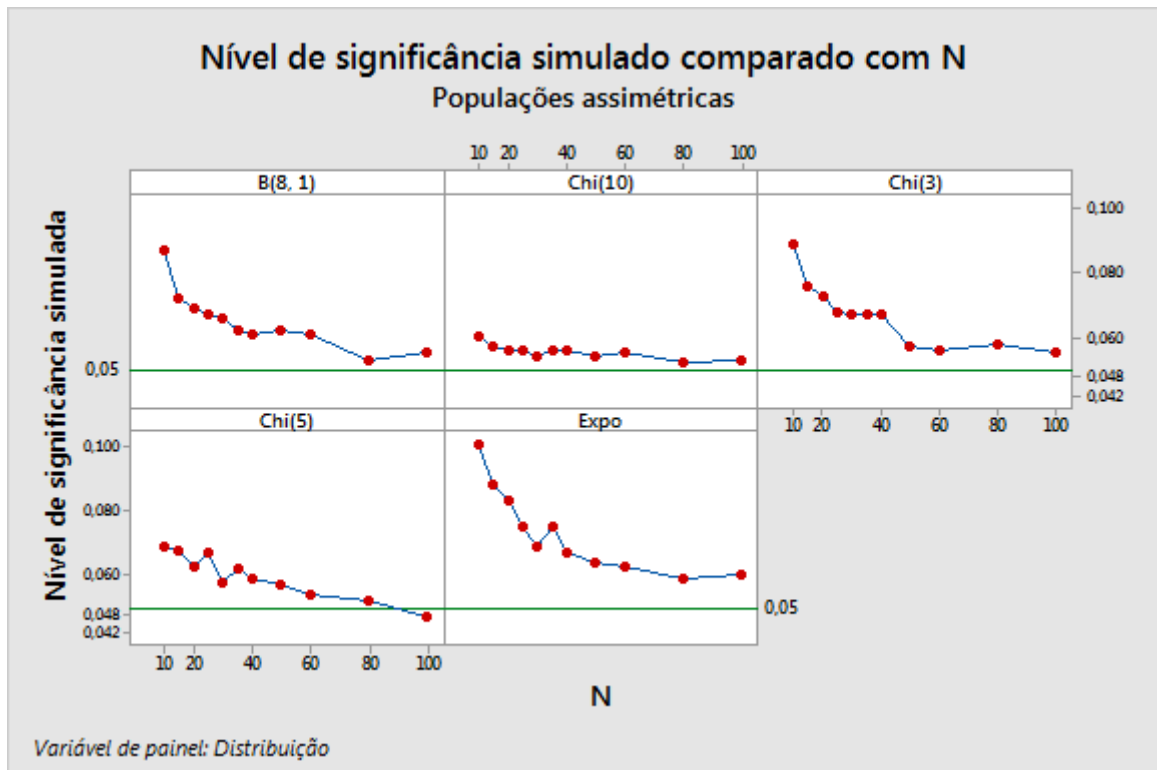


Figura 2 Gráficos de níveis de significância simulados de teste t para 1 amostra bilateral versus o tamanho de amostras geradas a partir de populações assimétricas. O nível de significância alvo é $\alpha = 0,05$.

Nessa investigação, focamos nos testes de hipóteses em vez de nos intervalos de confiança. No entanto, os resultados se estendem naturalmente aos intervalos de confiança porque tanto os testes de hipóteses quanto os intervalos de confiança podem ser usados para determinar significância estatística.

Apêndice B: Poder e tamanho amostral do teste

Tencionávamos examinar a sensibilidade da função de poder com relação à suposição normal sob a qual ela é derivada. Observe que, se β é o erro de tipo II de um teste, $1 - \beta$ é o poder do teste. Como resultado, o tamanho de amostra planejado é determinado de modo que a taxa de erro de tipo II seja pequena ou que, de forma equilibrada, o nível de poder seja razoavelmente alto.

As funções de poder dos testes t são bem conhecidas e documentadas. Pearson e Hartley (1954) e Neyman, Iwazskiewicz e Kolodziejczyk (1935) fornecem gráficos e tabelas de funções de poder.

No caso de um teste t para 1 amostra α bilateral, é possível fornecer uma expressão matemática dessa função em termos de tamanho de amostra e a diferença δ entre a média verdadeira μ e a média hipotética μ_0 como

$$\pi(n, \delta) = 1 - F_{n-1, \lambda}(t_{n-1}^{\alpha/2}) + F_{n-1, \lambda}(-t_{n-1}^{\alpha/2})$$

onde $F_{d, \lambda}(\cdot)$ é a C.D.F da distribuição t não central com $d = n - 1$ graus de liberdade e parâmetro de não centralidade

$$\lambda = \frac{\delta\sqrt{n}}{\sigma}$$

e t_d^α denota o ponto de percentil superior 100α da distribuição t com d graus de liberdade.

Para alternativas unilaterais, o poder é fornecido como

$$\pi(n, \delta) = 1 - F_{n-1, \lambda}(t_{n-1}^\alpha)$$

para testes de hipótese nula contra $\mu > \mu_0$, e fornecido como

$$\pi(n, \delta) = F_{n-1, \lambda}(-t_{n-1}^\alpha)$$

nos testes de hipótese nula contra $\mu < \mu_0$.

Essas funções são derivadas sob a suposição de que os dados estejam distribuídos normalmente e de que o nível de significância do teste esteja determinado em algum valor α .

Estudo de simulação B

Planejamos esta simulação com o objetivo de avaliar o efeito da não-normalidade sobre a função de poder teórica do teste t para 1 amostra. Para avaliar o efeito da não-normalidade, nós comparamos níveis de poder simulados com os níveis de poder alvo que foram calculados por meio da função de poder teórica do teste.

Realizamos testes t bilaterais em $\alpha = 0,05$ nas amostras aleatórias de vários tamanhos ($n = 10, 15, 20, 25, 30, 35, 40, 50, 60, 80, 100$) geradas a partir das mesmas populações descritas no primeiro estudo de simulação (consulte o apêndice A).

Para cada uma das populações, a hipótese nula do teste é $\mu = \mu_0 - \delta$ e sua hipótese alternativa é $\neq \mu_0 - \delta$, onde μ_0 é definido como a média verdadeira da população e $\delta = \sigma/2$ (σ corresponde ao desvio padrão da população parental). Com isso, a diferença entre a média verdadeira e a média hipotética é 0; sendo assim, a decisão correta consiste em rejeitar a hipótese nula.

Para cada tamanho de amostra, são obtidas réplicas de 10.000 amostras de cada distribuição. Em todos os tamanhos de amostra, a fração de 10.000 réplicas para as quais a hipótese nula é rejeitada representa o nível de poder simulado do teste em um dado tamanho de amostra e a diferença δ . Observe que escolhemos esse valor de diferença em particular porque ele produz valores de poder relativamente pequenos quando os tamanhos de amostra são pequenos.

Além disso, os valores de poder teóricos correspondentes utilizados como valores de poder alvo são calculados na diferença δ e nos vários tamanhos de amostra para comparação com os valores de poder simulados.

Os resultados da simulação estão na tabela 3 e na tabela 4 e são exibidos graficamente na figura 3 e na figura 4.

Resultados e sumário

Os resultados confirmam que o poder do teste t para 1 amostra é, de modo geral, insensível à suposição normal quando o tamanho da amostra é suficientemente grande. Nas amostras geradas a partir de populações simétricas, os resultados na tabela 3 mostram que os níveis de poder simulados e de poder alvo estão próximos inclusive nas amostras pequenas. As curvas de poder correspondentes exibidas na figura 3 são praticamente indistinguíveis. Nas amostras geradas a partir de distribuições normais contaminadas, os valores de poder são um tanto conservadores para tamanhos de amostra pequenos a moderados. É possível que esse resultado derive do fato de que o nível de significância real do teste daquelas populações é um pouco mais alto do que o nível de significância alvo determinado α .

Tabela 3 Níveis de poder simulados na diferença $\delta = \sigma/2$ de um teste t para 1 amostra bilateral $\alpha = 0,05$ quando as amostras são geradas a partir de populações simétricas. Os níveis de poder simulados são comparados com os níveis de poder alvo teóricos derivados sob suposição de normalidade.

n	Poder alvo	N(0, 1)	t(5)	t(10)	Lpl	CN(0,9, 3)	CN(0,8, 3)	B(3, 3)	U(0, 1)
		Nível de poder simulado em $\delta = \sigma/2$ (Populações simétricas)							
10	0,293	0,299	0,334	0,311	0,357	0,361	0,385	0,28	0,269
15	0,438	0,438	0,48	0,45	0,491	0,512	0,511	0,423	0,421
20	0,565	0,57	0,603	0,578	0,60	0,629	0,623	0,557	0,548
25	0,67	0,674	0,695	0,68	0,691	0,712	0,70	0,665	0,67
30	0,754	0,756	0,77	0,756	0,767	0,768	0,765	0,754	0,75
35	0,82	0,819	0,827	0,815	0,82	0,819	0,812	0,822	0,818
40	0,869	0,87	0,871	0,868	0,862	0,869	0,868	0,875	0,867
50	0,934	0,933	0,929	0,93	0,929	0,923	0,925	0,932	0,94
60	0,968	0,967	0,963	0,965	0,964	0,955	0,955	0,968	0,971
80	0,993	0,993	0,989	0,992	0,991	0,988	0,989	0,994	0,994
100	0,999	0,998	0,996	0,998	0,999	0,998	0,996	0,999	0,999

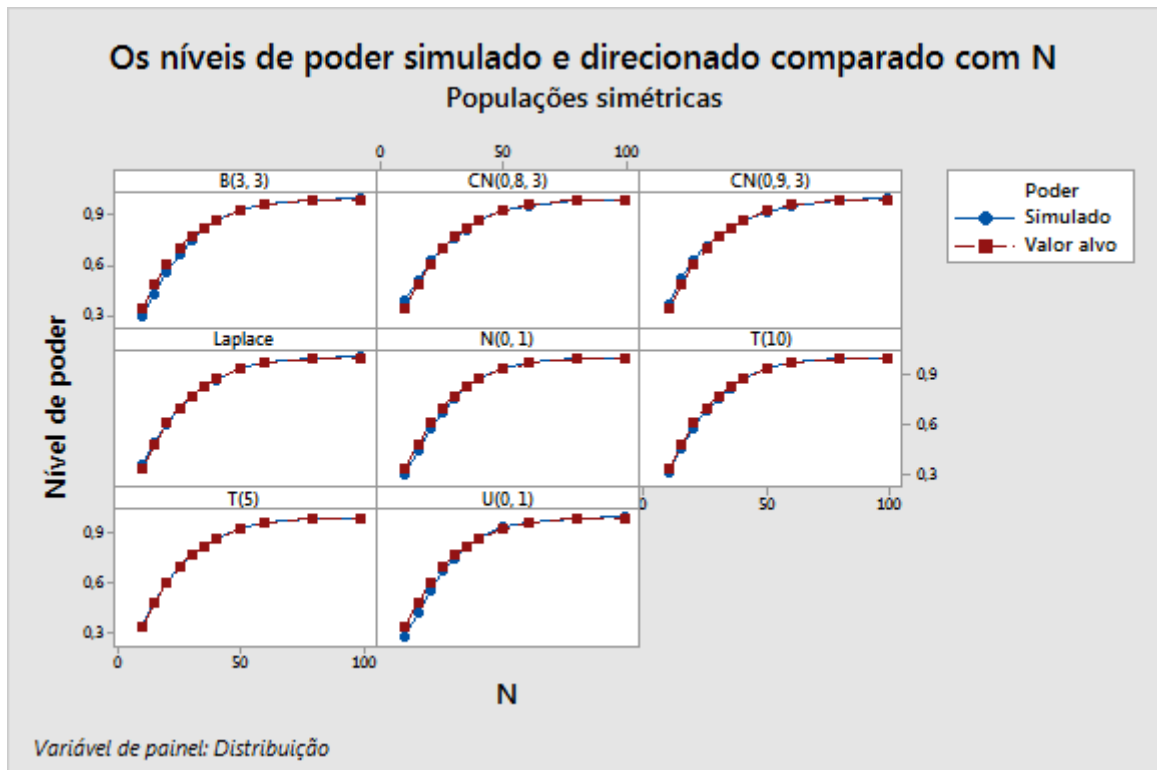


Figura 3 Curvas de poder simuladas comparadas a curvas de poder alvo teóricas em $\alpha = 0,05$, teste t para 1 amostra bilateral quando as amostras são geradas de populações simétricas. Os valores de poder são avaliados em um nível de diferença de $\delta = \sigma/2$.

Contudo, quando as amostras são obtidas de populações assimétricas, os valores de poder simulados estão fora do alvo de pequenas amostras, conforme mostrado na tabela 4 e na figura 4. Nas populações moderadamente assimétricas, como a distribuição Qui-quadrado com 5 graus de liberdade e Qui-quadrado com 10 graus de liberdade, quando o tamanho da amostra tem um valor mínimo de 20, os níveis de poder simulados e alvo se mantêm próximos. Por exemplo, em $n = 20$, o nível de poder alvo é 0,565 quando os níveis de poder são 0,576 e 0,577 para as distribuições Qui-quadrado 5 e Qui-quadrado 10, respectivamente. Nas distribuições extremamente assimétricas, são necessárias amostras maiores para que os níveis de poder simulados alcancem o nível de significância alvo. Isso acontece porque o teste t para 1 amostra não controla de forma apropriada o erro de tipo I quando os tamanhos de amostra são pequenos e as populações parentais são demasiadamente assimétricas.

Tabela 4 Valores de poder simulados na diferença $\delta = \sigma/2$ em teste t para 1 amostra bilateral $\alpha = 0,05$ quando as amostras são geradas de populações assimétricas. Os valores de poder simulados são comparados aos valores de poder alvo derivados sob a suposição de normalidade.

N	Poder alvo	Exp		Chi(3)	B(8, 1)	Chi(5)	Chi(10)
		Assimetria da população					
		2,0		1,633	-1,423	1,265	0,894
		Níveis de poder simulados					
10	0,293	0,206		0,212	0,39	0,225	0,238
15	0,438	0,416		0,413	0,484	0,409	0,407
20	0,565	0,604		0,591	0,566	0,576	0,577
25	0,67	0,763		0,734	0,657	0,709	0,695
30	0,754	0,859		0,834	0,729	0,808	0,785
35	0,82	0,917		0,895	0,776	0,874	0,835
40	0,869	0,955		0,935	0,823	0,925	0,905
50	0,934	0,987		0,981	0,90	0,973	0,96
60	0,968	0,997		0,994	0,937	0,991	0,985
80	0,993	1,00		0,999	0,98	0,999	0,997
100	0,999	1,00		1,00	0,994	1,00	1,00

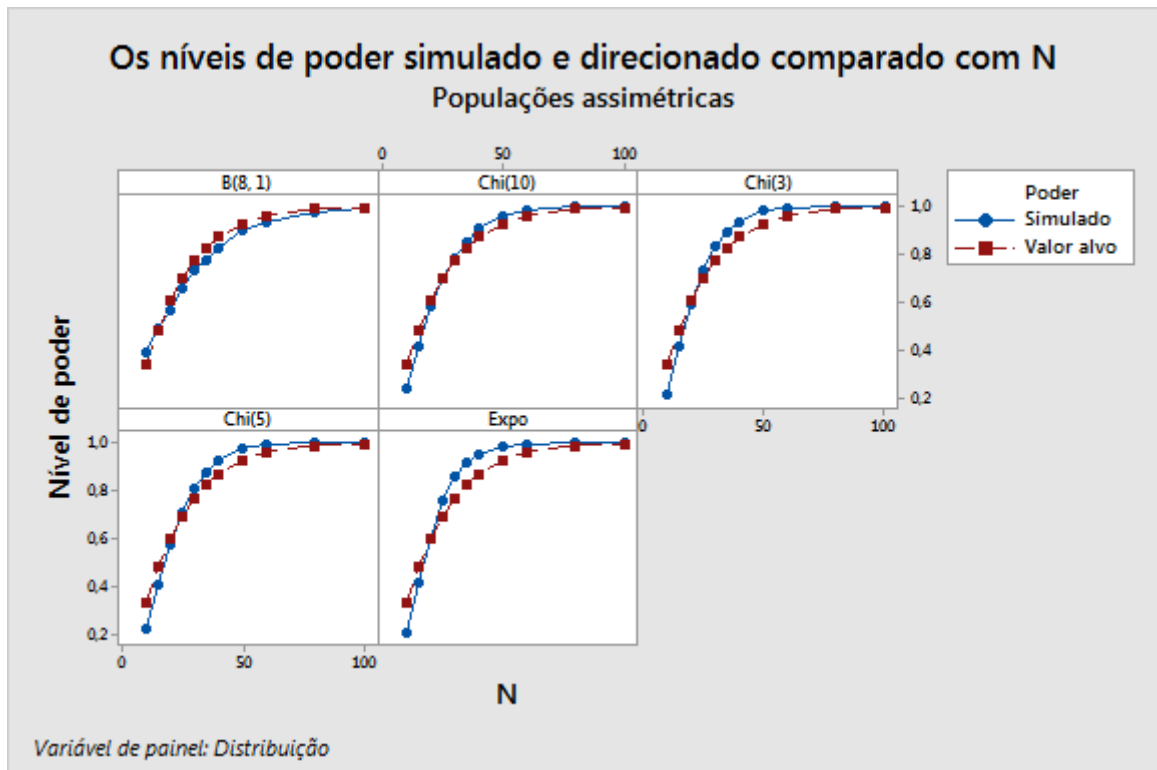


Figura 4 Curvas de poder simuladas comparadas a curvas de poder teóricas em $\alpha = 0,05$, teste t para 1 amostra bilateral quando as amostras são geradas de populações assimétricas. Os valores de poder são avaliados em uma diferença de $\delta = \sigma/2$.

Em resumo, nas distribuições moderadamente assimétricas, a função de poder será confiável se o tamanho de amostra tiver um valor mínimo de 20, independentemente da população parental da qual ela for obtida. Nas populações extremamente assimétricas, faz-se necessário um tamanho de amostra maior (aproximadamente 40) para que o poder simulado mantenha-se próximo ao poder alvo.