Este documento é de uma série de papéis que explicam a pesquisa conduzida por estatísticos da Minitab para desenvolver os métodos e as verificações de dados usadas no Assistente no Minitab Statistical Software.

Teste de % de defeituosos para 1 amostra

Visão geral

Um teste de 1 proporção usado para determinar se uma proporção difere de um valor alvo. Na análise de qualidade, o teste é frequentemente usado quando um produto ou serviço é caracterizado como defeituoso ou não defeituoso para determinar se o percentual de itens defeituosos difere significativamente de um alvo de % de defeituosos.

O Assistente do Minitab inclui um Teste de % de defeituosos para 1 amostra. Os dados coletados para o teste são o número de itens defeituosos em uma amostra, que se supõe que seja o valor observado de uma variável aleatória binomial. O Assistente usa os métodos exatos para calcular os resultados do teste de hipóteses e os intervalos de confiança; portanto, a taxa de erros do Tipo I atual deve estar próxima do nível de significância (alfa) especificado para o teste e nenhuma investigação adicional é necessária. Contudo, a potência e a análise do tamanho amostral do teste de % de defeituosos para 1 amostra são baseadas em uma aproximação e precisamos avaliá-la quanto à exatidão.

Neste documento investigamos a metodologia usada para avaliar o poder e o tamanho amostral do teste de % de defeituosos para 1 amostra, comparando o poder teórico do método aproximado com o poder real do teste exato.

Também descrevemos como estabelecimos um guia para ajudar você a avaliar se seu tamanho de amostra é grande o suficiente para detectar se o percentual dos itens defeituosos difere de um alvo % de defeituosos. O Assistente automaticamente realiza uma verificação no tamanho amostral e relata os achados no Cartão de Relatórios.

O Teste de % de defeituosos para 1 amostra também depende de outras suposições. Consulte o Apêndice A para obter detalhes.

Método de % de defeituosos para 1 amostra

Desempenho da função de poder teórico

O Assistente desempenha o teste de hipóteses para uma proporção única da população de Bernoulli (% de defeituosos) usando métodos exatos (razão de verossimilhança). Entretanto, como a função de poder deste teste exato não é facilmente derivada, a função de poder é aproximada usando-se a função de poder teórico do teste de aproximação normal correspondente.

Objetivo

Queremos determinar se podemos usar a função de poder teórico com base no teste de aproximação normal para avaliar o poder e os requisitos de tamanho de amostra do teste de % de defeituosos para 1 amostra no Assistente. Para fazer isso, precisamos avaliar se esta função de poder teórico reflete exatamente o poder real do teste (razão de verossimilhança) exato.

Método

A estatística do teste, valor-p e o intervalo de confiança do teste (razão de verossimilhança) exato são definidos no Apêndice B. A função de poder teórico com base na aproximação normal está definida no Apêndice C. Com base nessas definições, realizamos simulações para estimar os níveis de poder real (que nos referimos como níveis de poder simulado) usando o teste exato.

Para realizar as simulações, geramos amostras aleatórias de diversos tamanhos de diversas populações Bernoulli. Para cada população Bernoulli, realizamos o teste exato em cada uma das 10.000 réplicas de amostras. Para cada tamanho de amostra, calculamos o poder simulado do teste para detectar uma dada diferença como a fração das 10.000 amostras para as quais o teste é significativo. Para comparação, também calculamos o poder teórico correspondente com base no teste de aproximação normal. Se a aproximação funcionar bem, os níveis de poder teórico e simulado devem ser próximos. Para obter mais detalhes, consulte o Apêndice D.

Resultados

Nossas simulações mostraram que, em geral, a função de poder teórico do teste de aproximação normal e a função de poder simulado do teste (razão de verossimilhança) exato são praticamente iguais. Portanto, o Assistente usa a função de poder teórico do teste de aproximação normal para estimar os tamanhos das amostras necessários para assegurar que o teste exato tem poder suficiente para detectar diferenças praticamente importantes no percentual de defeituosos.

Verificações de dados

Tamanho amostral

Tipicamente, um teste de hipótese é realizado para coletar evidências para rejeitar a hipótese nula de “nenhuma diferença”. Se a amostra é muito pequena, o poder do teste pode não ser adequado para detectar uma diferença que realmente existe, que resulta em um erro do Tipo II. Portanto, é crucial assegurar que os tamanhos amostrais sejam suficientemente grandes para detectar diferenças praticamente importantes com alta probabilidade.

Objetivo

Se os dados não fornecerem evidência suficiente para rejeitar a hipótese nula, queríamos determinar se os tamanhos amostrais são grandes o suficiente para o teste para detectar diferenças práticas de interesse com alta probabilidade. Apesar de o objetivo do planejamento de tamanho amostral ser assegurar que os tamanhos amostrais sejam grandes o suficiente para detectar diferenças importantes com alta probabilidade, elas não devem ser tão grandes que diferenças inexpressivas tornem-se estatisticamente significativas com alta probabilidade.

Método

O poder e a análise do tamanho amostral do teste de % de defeituosos para 1 amostra está baseado na função de poder teórico que usa a aproximação normal, o que fornece uma boa estimativa do poder real do teste exato (consulte a seção do método de % de defeituosos para 1amostra acima). Quando o alvo de % de defeituosos é dado, a função de poder teórico depende do tamanho amostral e a diferença que você deseja detectar.

Resultados

Quando os dados não fornecem evidência suficiente contra a hipótese nula, o Assistente calcula as diferenças práticas que podem ser detectadas com uma probabilidade de 80% e de 90% para o tamanho amostral dado. Além disso, se o usuário fornecer uma diferença prática particular de interesse, o Assistente calcula tamanhos amostrais que produzem uma chance de 80% e de 90% de detecção da diferença.

Para ajudar a interpretar os resultados, o Cartão de Relatórios do Assistente do teste de % de defeituosos para 1 amostra exibe os seguintes indicadores de status ao verificar o poder e o tamanho da amostra:

Status	Condição
	O teste encontra uma diferença entre o % de defeituosos e o valor alvo, portanto, o poder não é um problema. OU O poder é suficiente. O teste não encontrou uma diferença do valor alvo, mas a amostra é grande o suficiente para fornecer pelo menos uma chance de 90% de detecção da diferença dada (poder ≥ 0,90).
	O poder pode ser suficiente. O teste não encontrou uma diferença do valor alvo, mas a amostra é grande o suficiente para fornecer uma chance de 80% a 90% de detecção da diferença dada (0,80 ≤ poder < 0,90). O tamanho da amostra necessário para alcançar o poder de 90% é indicado.
	O poder pode não ser suficiente. O teste não encontrou uma diferença do valor alvo, e a amostra é grande o suficiente para fornecer uma chance de 60% a 80% de detecção da diferença dada (0,60 ≤ poder < 0,80). Os tamanhos amostrais necessários para alcançar o poder de 80% e o poder de 90% estão indicados.
	O poder não é suficiente. O teste não encontrou uma diferença do valor alvo, e a amostra não é grande o suficiente para fornecer pelo menos uma chance de 60% de detecção da diferença dada (poder < 0,60). Os tamanhos amostrais necessários para alcançar o poder de 80% e o poder de 90% estão indicados.
	O teste não encontrou uma diferença do valor alvo. Você não especificou uma diferença prática para detecção; portanto, o relatório indica as diferenças que você pôde detectar com uma chance de 80% e de 90%, com base no seu alfa e tamanho amostral.

Referências

Arnold, S.F. (1990). *Mathematical statistics*. Englewood Cliffs, NJ: Prentice Hall, Inc.

Casella, G., & Berger, R.L. (1990). *Statistical inference*. Pacific Grove, CA: Wadsworth, Inc.

Apêndice A: Suposições adicionais para % de defeituosos para 1 amostra

O teste de % de defeituosos para 1 amostra está baseado nas seguintes suposições:

- Os dados consistem em n itens distintos com cada item classificado como defeituoso ou não defeituoso.
- A probabilidade de um item ser defeituoso é a mesma para cada item dentro de uma amostra.
- A verossimilhança de um item ser defeituoso não é afetada por se outro item é defeituoso ou não.

Essas suposições não podem ser verificadas nas verificações de dados do Cartão de Relatórios porque os dados do resumo, em vez dos dados brutos, são inseridos para este teste.

Apêndice B: teste (razão de verossimilhança) exato

Suponha que nós observamos uma amostra aleatória de $X_1, \dots, X_n$ de uma distribuição Bernoulli com probabilidade de sucesso $p = \Pr(X_i = 1) = 1 - \Pr(X_i = 0)$.

Os métodos exatos para traçar uma inferência sobre p estão descritos a seguir.

Fórmula B1: Teste exato e valor-p

Considere um teste das hipóteses nulas de $H_0: p = p_0$ contra qualquer dessas hipóteses alternativas: $H_A: p > p_0$, $H_A: p < p_0$ ou $H_A: p \neq p_0$.

Permita $X = \sum_{i=1}^{n} X_i$

Então, X é uma variável aleatória binomial com número de ensaios n e probabilidade de sucesso p.

Um teste unilateral baseado em X é UMP (uniformemente mais poderoso) e um teste de razão de verossimilhança. Para testes bilaterais, o teste da razão de verossimilhança também está baseado em X e a estatística do teste é

$$\Lambda(X) = \left(\frac{\hat{p}}{p_0}\right)^X \left(\frac{1-\hat{p}}{1-p_0}\right)^{n-X}$$

(consulte Arnold, 1990).

Os valores-p para teste unilaterais podem ser diretamente obtidos com base na distribuição exata de X. Para testes bilaterais, os valores-p são calculados como a probabilidade, sob a hipótese nula, de observação de uma razão de verossimilhança (ou razão de verossimilhança de log), no mínimo, tão grande quanto aquela observada realmente. Um algoritmo de achado de raiz numérica é geralmente usado para calcular esta probabilidade.

Fórmula B2: intervalo de confiança exato

Um intervalo de confiança bilateral de $100(1-\alpha)\%$ para p é

$$\frac{1}{1 + \frac{n-x+1}{x} F_{2(n-x+1),2x,\alpha/2}} \leq p \leq \frac{\frac{x+1}{n-x} F_{2(x+1),2(n-x),\alpha/2}}{1 + \frac{x+1}{n-x} F_{2(x+1),2(n-x),\alpha/2}}$$

onde x é o número observado de sucessos e $F_{\nu_1,\nu_2,\alpha}$ é o ponto percentil superior de α da distribuição de F com ν_1 e ν_2 graus de liberdade (consulte Casella e Berger, 1990). Adotamos a convenção de que o limite inferior é de 0 se $x = 0$ e o limite superior é 1 se $x = n$.

Apêndice C: função de poder teórico

Uma função de poder teórico do teste exato é muito complexa para derivar. Portanto, aproximamos a função de poder do teste usando a função de poder teórico dos teste com base na aproximação normal. Este teste aproximado está baseado no fato que de a variável aleatória

$$Z = \frac{n^{1/2}(\hat{p} - p)}{(p(1-p))^{1/2}}$$

é assintoticamente distribuída como a distribuição normal padrão. A função de poder teórico deste teste é bem conhecida e documentada. Para a hipótese alternativa bilateral, a função de poder é dada por:

$$\pi(n, \delta) = 1 - \Phi\left(\frac{-\delta + z_{\alpha/2}\sqrt{p_o(1-p_o)/n}}{\sqrt{p(1-p)/n}}\right) + \Phi\left(\frac{-\delta - z_{\alpha/2}\sqrt{p_o(1-p_o)/n}}{\sqrt{p(1-p)/n}}\right)$$

onde $p = \delta + p_o$, $\Phi(.)$ é a função da distribuição cumulativa da distribuição normal padrão, e z_α é o percentil superior da distribuição normal padrão.

Para a alternativa unilateral $H_A: p > p_0$ a função de poder pode ser dada como

$$\pi(n, \delta) = 1 - \Phi\left(\frac{-\delta + z_{\alpha}\sqrt{p_o(1-p_o)/n}}{\sqrt{p(1-p)/n}}\right)$$

Ao testar contra a alternativa unilateral $H_A: p < p_0$ a função de poder também pode ser dada como

$$\pi(n, \delta) = \Phi\left(\frac{-\delta - z_{\alpha}\sqrt{p_o(1-p_o)/n}}{\sqrt{p(1-p)/n}}\right)$$

Apêndice D: comparação do poder real versus o poder teórico

Simulação D1: estimando o poder real usando o teste exato

Planejamos uma simulação para comparar os níveis de poder real estimado (referidos como níveis de poder simulados) para os níveis de poder teóricos com base na função de poder do teste de aproximação normal (referido como níveis de poder aproximados). Em cada experimento, geramos 10.000 amostras, de cada tamanho n, a partir de uma população Bernoulli com uma probabilidade de sucesso p dada. Consideramos dois casos para a probabilidade de sucesso: (1) uma probabilidade moderada de sucesso, com o valor de p próximo de 0,5 (especificamente, $p = 0{,}45$) e (2) uma pequena ou grande probabilidade de sucesso, com o valor de p próximo de 0 ou 1 (especificamente, $p = 0{,}85$). Consideramos esses dois casos porque a aproximação normal DeMoivre-Laplace para a distribuição binomial, da qual o teste de aproximação normal é derivado, é conhecida por ser exata quando o tamanho da amostra Bernoulli é maior do que 10 e a probabilidade de sucesso é quase 0,5. Contudo, para probabilidades de sucesso maiores ou menores, são necessárias amostras Bernoulli maiores para aquela aproximação ser exata.

Em cada experimento, fixamos o tamanho da amostra em um valor único de n, onde $n = 10, 15, 20, 30, \ldots, 100$. Em todos os experimentos fixamos a diferença a ser detectada de $\delta = p - p_0$ em 0,2, para assegurar que os valores de poder resultantes não fossem muito pequenos ou muito grandes uma vez que o tamanho da amostra aumentou para 100. Para estimar o poder real para o teste baseado nos resultados de cada simulação, nós calculamos a fração das 10.000 réplicas de amostra para as quais o teste exato foi significativo no nível de significância alvo de $\alpha = 0{,}05$, usando ambos os testes exatos unilateral e bilateral. Por fim, calculamos os níveis de poder teóricos correspondentes com base no teste de aproximação normal para comparação. Os resultados estão indicados na Tabela 1 a seguir.

Tabela 1 Níveis de poder simulados e aproximados (aprox.) dos testes exatos bilateral e unilateral. O nível de significância alvo é $\alpha = 0{,}05$.

n	Teste bilateral				Teste unilateral			
	$p = 0,45$		$p = 0,85$		$p = 0,45$		$p = 0,85$	
	Poder simulado	Poder aprox.	Poder simulado	Poder aprox.	Poder simulado	Poder aprox.	Poder simulado	Poder aprox.
10	0,101	0,333	0,20	0,199	0,257	0,436	0,20	0,335
15	0,339	0,441	0,322	0,327	0,339	0,55	0,322	0,489
20	0,406	0,537	0,409	0,455	0,59	0,643	0,648	0,621

n	Teste bilateral				Teste unilateral			
	$p = 0,45$		$p = 0,85$		$p = 0,45$		$p = 0,85$	
	Poder simulado	Poder aprox.	Poder simulado	Poder aprox.	Poder simulado	Poder aprox.	Poder simulado	Poder aprox.
30	0,632	0,69	0,708	0,674	0,632	0,779	0,708	0,808
40	0,781	0,799	0,863	0,822	0,781	0,867	0,863	0,911
50	0,877	0,872	0,874	0,91	0,877	0,921	0,933	0,961
60	0,878	0,92	0,942	0,957	0,922	0,954	0,969	0,984
70	0,925	0,951	0,972	0,981	0,953	0,973	0,987	0,994
80	0,954	0,971	0,986	0,992	0,986	0,985	0,993	0,998
90	0,971	0,982	0,993	0,996	0,991	0,991	0,996	0,999
100	0,989	0,99	0,998	0,999	0,994	0,995	0,999	1,00

Os resultados mostram que os níveis de poder simulado e os níveis de poder aproximado são geralmente muito consistentes. Você pode ver esta consistência mais claramente quando os resultados são graficamente exibidos como curvas de poder, conforme mostrado nas Figuras 1 e 2 a seguir.

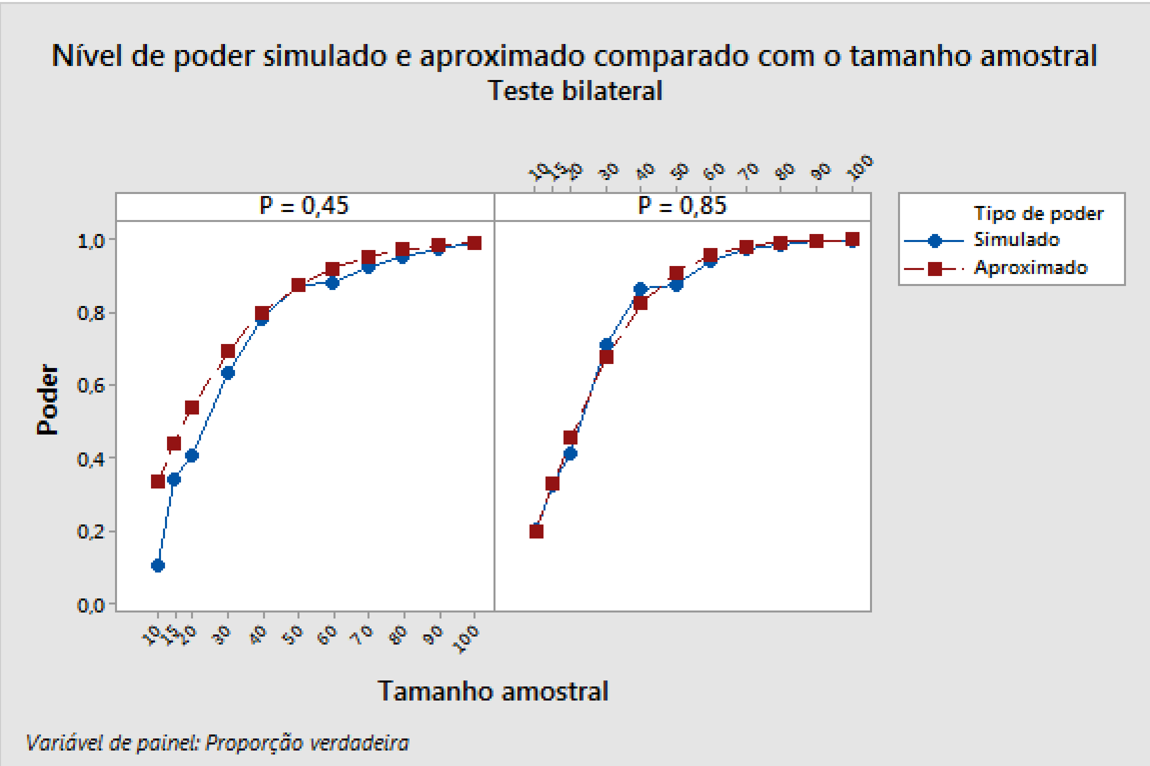

Figura 1 Gráficos de níveis de poder simulado e aproximado do teste exato bilateral comparados ao tamanho amostral.

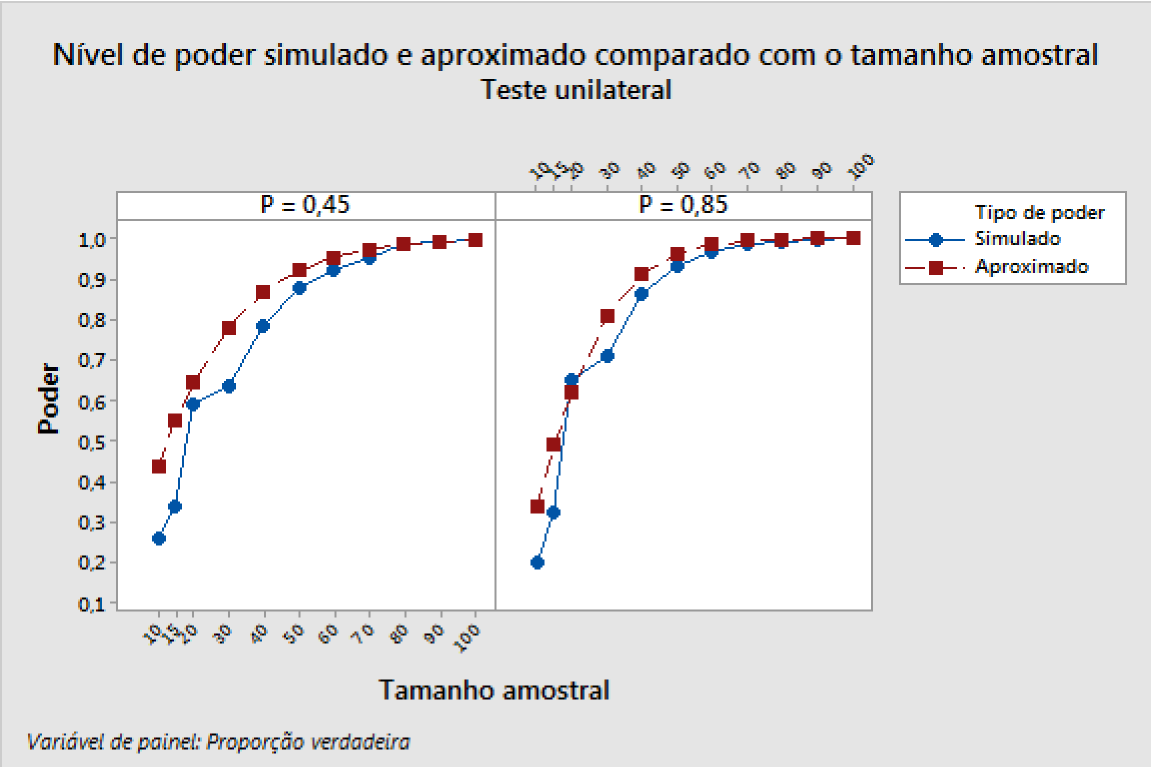

Figura 2 Gráficos de níveis de poder simulado e aproximado do teste exato unilateral comparado ao tamanho amostral.

As duas curvas de poder mostradas em cada painel nas Figuras 1 e 2 estão próximas uma da outra, exceto em algumas instâncias quando o tamanho da amostra é pequeno. A proximidade das curvas indica que a função de poder aproximado corresponde de pero ao poder simulado quando o teste exato é aplicado na prática. Portanto, é apropriado usar a função de poder aproximado para estimar o tamanho amostral.

As Figuras 1 e 2 também mostram que as curvas de poder teórico (aproximado) são em geral mais altas do que as curvas de poder simulado. As curvas de poder aproximado são mais altas porque os níveis de poder teórico são calculados supondo-se um valor exato para o nível de significância alvo (0,05). Em comparação, o teste exato tende a ser conservador, particularmente em pequenas amostras e, portanto, produz níveis de significância real que são menores do que o nível alvo. Como resultado, os níveis de poder simulado tendem a ser menores quando os tamanhos amostrais são pequenos.

Em conclusão, nossas simulações mostram que a função de poder teórico do teste de aproximação normal se aproximam de perto do poder do teste exato (razão de verossimilhança). Como resultado, a função de poder teórico do teste de aproximação normal fornece uma base sólida para estimar os tamanhos amostrais necessários para assegurar que o teste exato tem poder suficiente para detectar diferenças praticamente importantes.