

RESUMO DO ASSISTENTE DO MINITAB

Este artigo é parte de uma série de artigos que explicam a pesquisa conduzida pelos estatísticos do Minitab para desenvolver os métodos e verificações de dados usados no Assistente no Minitab Statistical Software.

Teste de % de defeituosos para 2 amostras

Visão geral

Um teste para 2 proporções é usado para determinar se duas proporções diferem significativamente. 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 para amostras coletadas de dois processos independentes.

O Assistente do Minitab inclui um Teste de % de defeituosos para 2 amostras. Os dados coletados para o teste são o número de itens defeituosos em cada uma das duas amostras independentes, 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; 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, o Assistente usa um método de aproximação normal para calcular o intervalo de confiança (CI) para a diferença em % de defeituosos e uma função de poder teórico do teste de aproximação normal para realizar suas análises de poder e de tamanho amostral. Devido a esses métodos de aproximação, precisamos avaliá-los quanto à exatidão.

Neste artigo, investigamos as condições sob as quais os intervalos de confiança aproximados são exatos. Também investigamos a metodologia usada para avaliar o poder e o tamanho amostral do Teste de % de defeituosos para 2 amostras, comparando o poder teórico do método aproximado com o poder real do teste exato. Por fim, examinamos as seguintes verificações de dados que são automaticamente realizadas e exibidas no Assistente de Cartão de Relatórios e explicamos como elas afetam os resultados da análise:

- Validade do CI
- Tamanho amostral

O Teste de % de defeituosos para 2 amostras também depende de outras suposições. Consulte o Apêndice A para obter detalhes.

Métodos do teste de % de defeituosos para 2 amostras

Exatidão dos intervalos de confiança

Apesar de o Assistente usar o teste exato de Fisher para avaliar se o % de defeituosos das duas amostras difere significativamente, o intervalo de confiança para a diferença está baseado no método de aproximação normal. De acordo com a regra geral encontrada na maioria dos livros de estatística, este intervalo de confiança aproximado é exato se o número observado de defeituosos e o número observado de não defeituosos em cada amostra for de, no mínimo, 5.

Objetivo

Queríamos avaliar as condições sob as quais os intervalos de confiança, baseados na aproximação normal, são exatos. Especificamente, queríamos ver como a regra geral relacionada ao número de defeituosos e de não defeituosos em cada amostra afeta a exatidão dos intervalos de confiança aproximados.

Método

A fórmula usada para calcular o intervalo de confiança, da diferença entre as duas proporções e a regra geral para assegurar sua exatidão, está descrita no Apêndice D. Além disso, descrevemos uma regra modificada e menos rigorosa que desenvolvemos durante o curso da nossa investigação.

Realizamos simulações para avaliar a exatidão do intervalo de confiança aproximado sob diversas condições. Para realizar as simulações, geramos pares de amostras aleatórias de diversos tamanhos de diversas populações Bernoulli. Para cada tipo de população Bernoulli, calculamos um intervalo de confiança aproximado para a diferença entre as duas proporções em cada par de 10.000 réplicas de amostras Bernoulli. Em seguida, calculamos a proporção dos 10.000 intervalos que contêm a verdadeira diferença entre as duas proporções, mencionadas como a probabilidade de cobertura simulada. Se o intervalo aproximado for exato, a probabilidade de cobertura simulada deverá próxima da probabilidade de cobertura do alvo de 0,95. Para avaliar a exatidão do intervalo aproximado em relação às regras originais e modificadas para o número mínimo de defeituosos e de não defeituosos exigido em cada amostra, também calculamos o percentual dos 10.000 pares de amostras para os quais cada regra foi satisfeita. Para obter mais detalhes, consulte o Apêndice D.

Resultados

O intervalo de confiança aproximado para a diferença entre duas proporções é geralmente exato quando as amostras são suficientemente grandes — isto é, quando o número observado de defeituosos e o número observado de não defeituosos em cada amostra é de, no mínimo, 5. Portanto, adotamos esta regra para nossa verificação de Validade de CI no Cartão de Relatórios. Apesar de esta regra geralmente apresentar bom desempenho, em

alguns casos ela pode ser excessivamente conservadora, e ela pode ser um pouco relaxada quando as duas proporções estão próximas de 0 ou 1. Para obter mais detalhes, consulte a seção Verificação de dados e o Apêndice D.

Desempenho da função de poder teórico

O Assistente realiza o teste de hipóteses para comparar duas proporções da população Bernoulli (% de defeituosos em duas amostras) usando-se o teste de Fisher. Entretanto, como a função de poder deste teste exato não é facilmente derivada, a função de poder deve ser aproximada usando-se a função de poder teórico do teste de aproximação normal correspondente.

Objetivo

Queríamos determinar se a função de poder teórico com base no teste de aproximação normal é apropriada para ser usada para avaliar o poder e os requisitos de tamanho amostral do teste de % de defeituosos para 2 amostras no Assistente. Para fazer isso, precisamos avaliar se esta função de poder teórico reflete exatamente o poder real do teste exato de Fisher.

Método

A metodologia do teste exato de Fisher, incluindo o cálculo de seu valor-p, está descrita em detalhes no Apêndice B. A função de poder teórico com base no teste de 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) do teste exato de Fisher quando ele é usado para analisar a diferença em % de defeituosos de duas amostras.

Para realizar as simulações, geramos pares de amostras aleatórias de diversos tamanhos de diversas populações Bernoulli. Para cada categoria da população Bernoulli, realizamos o teste exato de Fisher em cada uma das 10.000 réplicas de amostras. Para cada tamanho amostral, calculamos o poder simulado do teste para detectar uma dada diferença como a fração dos 10.000 pares de amostras para as quais o teste foi 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 E.

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 exato de Fisher são praticamente iguais. Portanto, o Assistente usa a função de poder teórico do teste de aproximação normal para estimar os tamanhos amostrais necessários para detectar diferenças praticamente importantes ao realizar o teste exato de Fisher.

Verificações dos dados

Validade do CI

Como o teste % de defeituosos para 2 amostras usa um teste exato para avaliar a diferença em % de defeituosos, sua exatidão não é muito afetada pelo número de defeituosos e de não defeituosos em cada amostra. Contudo, o intervalo de confiança da diferença entre o % de defeituosos é baseado em uma aproximação normal. Quando o número de defeituosos e de não defeituosos em cada amostra aumenta, a exatidão do intervalo de confiança aproximado também aumenta (consulte o Apêndice D).

Objetivo



Queríamos determinar se o número de defeituosos e o número de não defeituosos nas amostras são suficientes para assegurar que o intervalo de confiança aproximado para a diferença em % de defeituosos é exato.

Método

Usamos a regra geral encontrada na maioria dos livros de estatística. Quando cada amostra contém, no mínimo, 5 defeituosos e 5 não defeituosos, o intervalo de confiança aproximado para o teste de % de defeituosos para 2 amostras é exato. Para obter mais detalhes, consulte a seção de métodos de % de defeituosos para 2 amostras acima.

Resultados

Conforme mostrado nas simulações resumidas na seção Método de % de defeituosos para 2 amostras, a exatidão do intervalo de confiança depende do número mínimo de defeituosos e não defeituosos em cada amostra. Portanto, o Cartão de Relatórios do Assistente exibe os seguintes indicadores de status para ajudá-lo a avaliar a exatidão do intervalo de confiança para a diferença entre dois % de defeituosos:

Status	Condição
	Ambas as amostras têm pelo menos 5 defeituosos e 5 não defeituosos. O intervalo de confiança da diferença deve ser exato.
	O número de defeituosos ou o número de não defeituosos em pelo menos uma amostra é menor do que 5. O intervalo de confiança da diferença pode não ser exato.

Tamanho amostral

Geralmente, um teste de hipótese estatístico 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, queremos 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 defeituosos para 2 amostras estão baseados na função de poder teórico do teste de aproximação normal, o que fornece uma boa estimativa do poder real do teste exato de Fisher (consulte os resultados da simulação resumidos em Desempenho da função do poder teórico na seção Método de % de defeituosos para 2 amostras). A função de poder teórico pode ser expressa como uma função da diferença alvo no % de defeituosos e no % de defeituosos geral nas amostras combinadas.

Resultados

Quando os dados não fornecem evidências suficientes contra a hipótese nula, o Assistente usa a função de poder do teste de aproximação normal para calcular as diferenças práticas, que podem ser detectadas com uma probabilidade de 80% e de 90% para o tamanho amostral fornecido. Além disso, se o usuário fornecer uma determinada diferença prática de interesse, o Assistente usa a função de poder do teste de aproximação normal para calcular os tamanhos amostrais que produzem 80% e 90% de chance de detecção da diferença.

Para ajudar a interpretar os resultados, o Cartão de Relatórios do Assistente do teste % de defeituosos para 2 amostras exibe os seguintes indicadores de status ao verificar o poder e o tamanho amostrais:

Status	Condição
	O teste encontra uma diferença entre os % de defeituosos, de modo que o poder não é um problema. OU O poder é suficiente. O teste não encontrou nenhuma diferença entre o % de defeituosos, mas a amostra é grande o suficiente para fornecer pelo menos 90% de chance de detectar a determinada diferença (poder $\geq 0,90$).
	O poder pode ser suficiente. O teste não encontrou uma diferença entre o % de defeituosos, mas a amostra é grande o suficiente para fornecer pelo menos 80% a 90% de chance de detectar a determinada diferença ($0,80 \leq \text{poder} < 0,90$). O tamanho amostral necessário para atingir 90% de poder é informado.

Status	Condição
	<p>Talvez o poder não seja suficiente. O teste não encontrou uma diferença entre o % de defeituosos, e a amostra é grande o suficiente para fornecer pelo menos 60% a 80% de chance de detectar a determinada diferença ($0,60 \leq \text{poder} < 0,80$). Os tamanhos amostrais necessários para atingir 80% e 90% de poder são informados.</p>
	<p>O poder não é suficiente. O teste não encontrou uma diferença entre o % de defeituosos, e a amostra não é grande o suficiente para fornecer pelo menos 60% de chance de detectar a determinada diferença ($\text{poder} < 0,60$). Os tamanhos amostrais necessários para atingir 80% e 90% de poder são informados.</p>
	<p>O teste não encontrou uma diferença entre o % de defeituosos. Você não especificou uma diferença prática para detectar. Dependendo dos seus dados, o relatório pode indicar as diferenças que você poderia detectar com 80% e 90% de chance, com base no seu alfa e tamanho amostral.</p>

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 teste de % de defeituosos para 2 amostras

O teste % de defeituosos para 2 amostras está baseado nas seguintes suposições:

- Os dados em cada amostra 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 se outro item é defeituoso ou não.

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

Apêndice B: Teste exato de Fisher

Suponha que observamos duas amostras aleatórias independentes X_1, \dots, X_{n_1} e Y_1, \dots, Y_{n_2} a partir de distribuições de Bernoulli, tais como

$$p_1 = \Pr(X_i = 1) = 1 - \Pr(X_i = 0) \text{ e } p_2 = \Pr(Y_j = 1) = 1 - \Pr(Y_j = 0)$$

Nas seções a seguir, descrevemos os procedimentos para fazer inferências sobre a diferença entre as proporções $\delta = p_1 - p_2$.

Fórmula B1: Teste exato de Fisher e valor-p

Uma descrição do teste exato de Fisher pode ser encontrada em Arnold (1994). Fornecemos uma breve descrição do teste.

Permita que V seja o número de sucessos na primeira amostra e que $v = n_1 \hat{p}_1$ seja o número observado de sucessos na primeira amostra quando um experimento é realizado. Além disso, permita que W seja o número total de sucessos nas duas amostras e que $w = n_1 \hat{p}_1 + n_2 \hat{p}_2$ seja os sucessos observados quando um experimento é realizado. Observe que \hat{p}_1 e \hat{p}_2 são as estimativas de ponto de amostras de p_1 e p_2 .

Sob a hipótese nula de que $\delta = p_1 - p_2 = 0$, a distribuição condicional de V fornecido W é a distribuição hipergeométrica com a função de massa de probabilidade

$$f(v|w) = \frac{\binom{n_1}{v} \binom{n_2}{w-v}}{\binom{n_1+n_2}{w}}$$

Permita que $F(v|w)$ seja a c.d.f da distribuição. Em seguida, os valores-p para os testes unilaterais ou bilaterais são:

- **Quando testados contra $\delta < 0$ ou equivalentemente $p_1 < p_2$**

O valor-p é calculado como $F(v|w)$, em que v é o valor observado de V ou o número observado de sucessos na primeira amostra e w é o valor observado de W ou o número observado de sucessos em ambas as amostras.

- **Quando testados contra $\delta > 0$ ou equivalentemente $p_1 > p_2$**

O valor-p é calculado como $1 - F(v-1|w)$, em que v é o valor observado de V ou o número observado de sucessos na primeira amostra e w é o valor observado de W ou o número observado de sucessos em ambas as amostras.

- Quando testados contra $\delta \neq 0$ ou equivalentemente $p_1 \neq p_2$

O valor-p é calculado de acordo com algoritmo a seguir, em que m é o modo de distribuição hipergeométrica descrito acima.

- Se $v < m$, então o valor-p é calculado como $1 - F(y - 1|w) + F(v|w)$, em que v e w são como definidos acima e $y = \min\{k \geq m: f(k|w) \leq f(v|W)\}$
- Se $v = m$, então o valor-p é 1,0
- Se $v > m$, então o valor-p é calculado como $1 - F(v - 1|w) + F(y|w)$, onde v e w são como definidos acima e $y = \max\{k \leq m: f(k|w) \leq f(v|W)\}$

Apêndice C: Função de poder teórico

Para comparar duas proporções (ou mais especificamente dois % de defeituosos), usamos o teste exato de Fisher, como descrito no Apêndice B. Como uma função de poder teórico deste teste é muito complexa para derivar, usamos uma função de poder aproximado. Mais especificamente, usamos a função de poder do teste de aproximação normal bem conhecido para duas proporções, para aproximar o poder do teste exato de Fisher.

A função de poder da aproximação normal para o teste bilateral é

$$\pi(n_1, n_2, \delta) = 1 - \Phi\left(\frac{-\delta + z_{\alpha/2}\sqrt{p_c(1-p_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}{se}\right) + \Phi\left(\frac{-\delta - z_{\alpha/2}\sqrt{p_c(1-p_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}{se}\right)$$

em que $\delta = p_1 - p_2$,

$$se = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

e $p_c = (n_1p_1 + n_2p_2)/(n_1 + n_2)$.

Ao testar $p_1 = p_2$ contra $p_1 > p_2$ a função de poder é

$$\pi(n_1, n_2, \delta) = 1 - \Phi\left(\frac{-\delta + z_{\alpha}\sqrt{p_c(1-p_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}{se}\right)$$

Ao testar $p_1 = p_2$ contra $p_1 < p_2$ a função de poder é

$$\pi(n_1, n_2, \delta) = \Phi\left(\frac{-\delta - z_{\alpha}\sqrt{p_c(1-p_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}{se}\right)$$

Apêndice D: Intervalos de confiança aproximados

Fórmula D1: Calculando um intervalo de confiança aproximado para a diferença entre 2 proporções

Fórmula D1: Calculando um intervalo de confiança aproximado para a diferença entre 2 proporções

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} \sqrt{\hat{p}_1(1 - \hat{p}_1)/n_1 + \hat{p}_2(1 - \hat{p}_2)/n_2}$$

Uma regra geral bem conhecida para avaliar a confiabilidade deste intervalo de confiança aproximado é $n_1\hat{p}_1 \geq 5$, $n_1(1 - \hat{p}_1) \geq 5$, $n_2\hat{p}_2 \geq 5$ e $n_2(1 - \hat{p}_2) \geq 5$. Em outras palavras, o intervalo de confiança é exato se o número observado de sucessos e falhas em cada amostra é de, no mínimo, 5.

Observação: Nesta seção e nas seções a seguir, expressamos a regra para o intervalo de confiança em sua forma mais geral, nos termos do número de sucessos e no número de falhas de cada amostra. Um sucesso é o evento de interesse e uma falha é o complemento do evento de interesse. Portanto, no contexto específico do teste % de defeituosos para 2 amostras, o número de "sucessos" é equivalente ao número de defeituosos e o número de "falhas" é equivalente ao número de não defeituosos.

Fórmula D2: Regras para intervalos de confiança aproximado

A regra geral usada para intervalos de confiança com base na aproximação normal declara que os intervalos de confiança são exatos se $n_1\hat{p}_1 \geq 5$, $n_1(1 - \hat{p}_1) \geq 5$, $n_2\hat{p}_2 \geq 5$ e $n_2(1 - \hat{p}_2) \geq 5$. Isto é, o nível de confiança real do intervalo é igual a ou aproximadamente igual ao nível de confiança alvo se cada amostra contiver, no mínimo, 5 sucessos (defeituosas) e 5 falhas (não defeituosas).

A regra é expressa em termos das proporções estimadas de sucessos e falhas em oposição às proporções verdadeiras porque na prática as proporções verdadeiras são desconhecidas. Contudo, nas configurações teóricas onde as proporções reais são supostas ou conhecidas, a regra pode ser diretamente expressa em termos das proporções reais. Nesses casos, podemos avaliar diretamente como o número esperado real de sucessos e o número esperado de falhas, n_1p_1 , n_2p_2 , $n_1(1 - p_1)$ e $n_2(1 - p_2)$, afetam a probabilidade de cobertura real do intervalo de confiança da diferença entre as proporções.

Podemos avaliar a probabilidade de cobertura real com amostras de um grande número de pares de amostras de tamanhos n_1 e n_2 de duas populações Bernoulli com probabilidade de sucessos de p_1 e p_2 . A probabilidade de cobertura real é calculado como a frequência relativa dos pares de amostras que rendem intervalos de confiança que contêm a verdadeira diferença entre as duas proporções. Se a probabilidade da cobertura real for exata quando

$n_1p_1 \geq 5$, $n_2p_2 \geq 5$, $n_1(1 - p_1) \geq 5$ e $n_2(1 - p_2) \geq 5$ de acordo com a sólida lei dos grandes números, a probabilidade de cobertura é exata quando $n_1\hat{p}_1 \geq 5$, $n_1(1 - \hat{p}_1) \geq 5$, $n_2\hat{p}_2 \geq 5$ e $n_2(1 - \hat{p}_2) \geq 5$. Assim, quando o nível de confiança real e alvo estão próximos, deve-se esperar que uma grande proporção dos pares das amostras gerados a partir das duas populações Bernoulli sejam de tal forma que $n_1\hat{p}_1 \geq 5$, $n_1(1 - \hat{p}_1) \geq 5$, $n_2\hat{p}_2 \geq 5$ e $n_2(1 - \hat{p}_2) \geq 5$ se esta regra for válida. Na simulação que se segue, nos referimos a esta regra como Regra 1.

Além disso, no curso desta investigação, em muitos casos, observamos que se $n_1p_1 \geq 5$ e $n_2p_2 \geq 5$ ou se $n_1(1 - p_1) \geq 5$ e $n_2(1 - p_2) \geq 5$, a probabilidade de cobertura simulada do intervalo está próxima da cobertura do alvo. Isso solicitou uma regra alternativa e mais relaxada que declara que os intervalos de confiança aproximados são exatos se $n_1\hat{p}_1 \geq 5$ e $n_2\hat{p}_2 \geq 5$ ou se $n_1(1 - \hat{p}_1) \geq 5$ e $n_2(1 - \hat{p}_2) \geq 5$. Na simulação que se segue, nos referimos a esta regra modificada como Regra 2.

Simulação D1: Avaliando a exatidão de intervalos de confiança aproximados

Realizamos simulações para avaliar as condições sob as quais o intervalo de confiança aproximado para a diferença entre duas proporções é exato. Em particular, examinamos a exatidão do intervalo em relação às seguintes regras gerais:

Regra 1 (original) $n_1p_1 \geq 5$, $n_2p_2 \geq 5$, $n_1(1 - p_1) \geq 5$ e $n_2(1 - p_2) \geq 5$

Regra 2 (modificada) $n_1\hat{p}_1 \geq 5$ e $n_2\hat{p}_2 \geq 5$ OU $n_1(1 - \hat{p}_1) \geq 5$ e $n_2(1 - \hat{p}_2) \geq 5$

Em cada experimento, geramos 10.000 pares de amostras de pares de populações Bernoulli definidos pelas seguintes proporções:

- **Proporções A: ambos p_1 e p_2 estão perto de 1,0 (ou perto de 0).** Para representar este par de populações Bernoulli na simulação, usamos $p_1 = 0,8$ e $p_2 = 0,9$.
- **Proporções B: p_1 e p_2 estão próximos de 0,5.** Para representar este par de populações Bernoulli na simulação, usamos $p_1 = 0,4$ e $p_2 = 0,55$.
- **Proporções C: p_1 está perto de 0,5 e p_2 está perto de 1,0.** Para representar este par de populações Bernoulli na simulação, usamos $p_1 = 0,4$ e $p_2 = 0,9$.

A classificação de proporções acima é baseada na aproximação normal DeMoivre-Laplace à distribuição binomial da qual os intervalos de confiança aproximados são derivados. Esta aproximação normal é conhecida por ser exata quando a amostra Bernoulli é maior do que 10 e a probabilidade de sucesso é de quase 0,5. Quando a probabilidade de sucesso é de quase 0 ou 1, uma amostra Bernoulli maior é geralmente necessária.

Fixamos os tamanhos amostrais de ambos os pares em um valor único de n , em que $n = 10, 15, 20, 30, \dots, 100$. Limitamos o estudo a experimentos balanceados ($n_1 = n_2 = n$) sem nenhuma perda de generalidade devido a ambas as regras dependerem do número observado de sucessos e falhas, que podem ser controladas pelo tamanho das amostras e a proporção dos sucessos.

Para estimar o nível de confiança real do intervalo de confiança para a diferença nas proporções das duas populações (referido como nível de confiança simulado), calculamos a

proporção dos 10.000 intervalos que contêm a diferença real entre as duas proporções. A probabilidade da cobertura alvo em cada experimento foi de 0,95. Além disso, determinamos o percentual das 10.000 amostras para as quais as condições sob as duas regras foram satisfeitas.

Observação: Para algumas amostras pequenas, o erro padrão estimado da diferença entre as proporções foi 0. Consideramos essas amostras "degeneradas" e as descartamos do experimento. Como resultado, o número de réplicas de amostras foi ligeiramente menor do que 10.000 em alguns casos.

Os resultados são mostrados nas Tabelas 1-11 e graficamente exibidos na Figura 1 a seguir.

Tabela 1 Probabilidades de cobertura simulada e percentual de amostras que satisfazem a Regra 1 e a Regra 2 para $n=10$. A probabilidade de cobertura alvo é 0,95.

$n = 10$							
Categoria	Proporção (p)	np	$n(1 - p)$	Probabilidade de cobertura	% Amostras que satisfazem a Regra 1	% Amostras que satisfazem a Regra 2	
A	p_1	0,80	8,00	2,00	0,907	0,0	99,1
	p_2	0,90	9,00	1,00			
B	p_1	0,40	4,00	6,00	0,928	4,4	63,0
	p_2	0,55	5,50	4,50			
C	p_1	0,45	4,50	5,50	0,919	0,0	48,3
	p_2	0,90	9,00	1,00			

Tabela 2 Probabilidades de cobertura simulada e percentual de amostras que satisfazem a Regra 1 e a Regra 2 para $n=15$. A probabilidade de cobertura alvo é 0,95.

$n = 15$							
Categoria	Proporção (p)	np	$n(1 - p)$	Probabilidade de cobertura	% Amostras que satisfazem a Regra 1	% Amostras que satisfazem a Regra 2	
A	p_1	0,80	12,00	3,00	0,938	0,2	100,0
	p_2	0,90	13,50	1,50			
B	p_1	0,40	6,00	9,00	0,914	65,0	97,3
	p_2	0,55	8,25	6,75			
C	p_1	0,45	6,75	8,25	0,93	1,2	86,9
	p_2	0,90	13,50	1,50			

Tabela 3 Probabilidades de cobertura simulada e percentual de amostras que satisfazem a Regra 1 e a Regra 2 para $n=20$. A probabilidade de cobertura alvo é 0,95.

$n = 20$							
Categoria	Proporção (p)	np	$n(1 - p)$	Probabilidade e de cobertura	% Amostras que satisfazem a Regra 1	% Amostras que satisfazem a Regra 2	
A	p_1	0,80	16,00	4,00	0,942	1,5	100,0
	p_2	0,90	18,00	2,00			
B	p_1	0,40	8,00	12,00	0,943	92,8	99,9
	p_2	0,55	11,00	9,00			
C	p_1	0,45	9,00	11,00	0,934	4,1	98,2
	p_2	0,90	18,00	2,00			

Tabela 4 Probabilidades de cobertura simulada e percentual de amostras que satisfazem a Regra 1 e a Regra 2 para $n=30$. A probabilidade de cobertura alvo é 0,95.

$n = 30$							
Categoria	Proporção (p)	np	$n(1 - p)$	Probabilidade e de cobertura	% Amostras que satisfazem a Regra 1	% Amostras que satisfazem a Regra 2	
A	p_1	0,80	24,00	6,00	0,941	4,3	100,0
	p_2	0,90	27,00	3,00			
B	p_1	0,40	12,00	18,00	0,944	99,7	100,0
	p_2	0,55	16,50	13,50			
C	p_1	0,45	13,50	16,50	0,938	7,2	100,0
	p_2	0,90	27,00	3,00			

Tabela 5 Probabilidades de cobertura simulada e percentual de amostras que satisfazem a Regra 1 e a Regra 2 para $n=40$. A probabilidade de cobertura alvo é 0,95.

$n = 40$							
Categoria	Proporção (p)	np	n(1 - p)	Probabilidade e de cobertura	% Amostras que satisfazem a Regra 1	% Amostras que satisfazem a Regra 2	
A	p_1	0,80	32,00	8,00	0,941	35,1	100,0
	p_2	0,90	36,00	4,00			
B	p_1	0,40	16,00	24,00	0,945	100,0	100,0
	p_2	0,55	22,00	18,00			
C	p_1	0,45	18,00	22,00	0,945	37,7	100,0
	p_2	0,90	36,00	4,00			

Tabela 6 Probabilidades de cobertura simulada e percentual de amostras que satisfazem a Regra 1 e a Regra 2 para $n=50$. A probabilidade de cobertura alvo é 0,95.

$n = 50$							
Categoria	Proporção (p)	np	n(1 - p)	Probabilidade e de cobertura	% Amostras que satisfazem a Regra 1	% Amostras que satisfazem a Regra 2	
A	p_1	0,80	40,00	10,00	0,942	36,4	100,0
	p_2	0,90	45,00	5,00			
B	p_1	0,40	20,00	30,00	0,944	100,0	100,0
	p_2	0,55	27,50	22,50			
C	p_1	0,45	22,50	27,50	0,935	38,3	100,0
	p_2	0,90	45,00	5,00			

Tabela 7 Probabilidades de cobertura simulada e percentual de amostras que satisfazem a Regra 1 e a Regra 2 para $n=60$. A probabilidade de cobertura alvo é 0,95.

$n = 60$							
Categoria	Proporção (p)	np	$n(1 - p)$	Probabilidade e de cobertura	% Amostras que satisfazem a Regra 1	% Amostras que satisfazem a Regra 2	
A	p_1	0,80	48,00	12,00	0,947	72,8	100,0
	p_2	0,90	54,00	6,00			
B	p_1	0,40	24,00	36,00	0,947	100,0	100,0
	p_2	0,55	33,00	27,00			
C	p_1	0,45	27,00	33,00	0,949	73,1	100,0
	p_2	0,90	54,00	6,00			

Tabela 8 Probabilidades de cobertura simulada e percentual de amostras que satisfazem a Regra 1 e a Regra 2 para $n=70$. A probabilidade de cobertura alvo é 0,95.

$n = 70$							
Categoria	Proportion (p)	np	$n(1 - p)$	Probabilidade e de cobertura	% Amostras que satisfazem a Regra 1	% Amostras que satisfazem a Regra 2	
A	p_1	0,80	56,00	14,00	0,939	71,7	100,0
	p_2	0,90	63,00	7,00			
B	p_1	0,40	28,00	42,00	0,945	100,0	100,0
	p_2	0,55	38,50	31,50			
C	p_1	0,45	31,50	38,50	0,944	71,8	100,0
	p_2	0,90	63,00	7,00			

Tabela 9 Probabilidades de cobertura simulada e percentual de amostras que satisfazem a Regra 1 e a Regra 2 para $n=80$. A probabilidade de cobertura alvo é 0,95.

$n = 80$							
Categoria	Proporção (p)	np	$n(1 - p)$	Probabilidade e de cobertura	% Amostras que satisfazem a Regra 1	% Amostras que satisfazem a Regra 2	
A	p_1	0,80	64,00	16,00	0,947	91,3	100,0
	p_2	0,90	72,00	8,00			
B	p_1	0,40	32,00	48,00	0,947	100,0	100,0
	p_2	0,55	44,00	36,00			
C	p_1	0,45	36,00	44,00	0,948	91,3	100,0
	p_2	0,90	72,00	8,00			

Tabela 10 Probabilidades de cobertura simulada e percentual de amostras que satisfazem a Regra 1 e a Regra 2 para $n=90$. A probabilidade de cobertura alvo é 0,95.

$n = 90$							
Categoria	Proporção (p)	np	$n(1 - p)$	Probabilidade e de cobertura	% Amostras que satisfazem a Regra 1	% Amostras que satisfazem a Regra 2	
A	p_1	0,80	72,00	18,00	0,947	95,18	100,0
	p_2	0,90	81,00	9,00			
B	p_1	0,40	36,00	54,00	0,951	100,0	100,0
	p_2	0,55	49,50	40,50			
C	p_1	0,45	40,50	49,50	0,945	95,2	100,0
	p_2	0,90	81,00	9,00			

Tabela 11 Probabilidades de cobertura simulada e percentual de amostras que satisfazem a Regra 1 e a Regra 2 para $n=100$. A probabilidade de cobertura alvo é 0,95.

$n = 100$							
Categoria	Proporção (p)	np	$n(1 - p)$	Probabilidade e de cobertura	% Amostras que satisfazem a Regra 1	% Amostras que satisfazem a Regra 2	
A	p_1	0,80	80,00	20,00	0,952	97,7	100,0
	p_2	0,90	90,00	10,00			

n = 100							
Categoria		Proporção (p)	np	n(1 - p)	Probabilidade e de cobertura	% Amostras que satisfazem a Regra 1	% Amostras que satisfazem a Regra 2
B	p ₁	0,40	40,00	60,00	0,945	100,0	100,0
	p ₂	0,55	55,00	45,00			
C	p ₁	0,45	45,00	55,00	0,948	97,7	100,0
	p ₂	0,90	90,00	10,00			

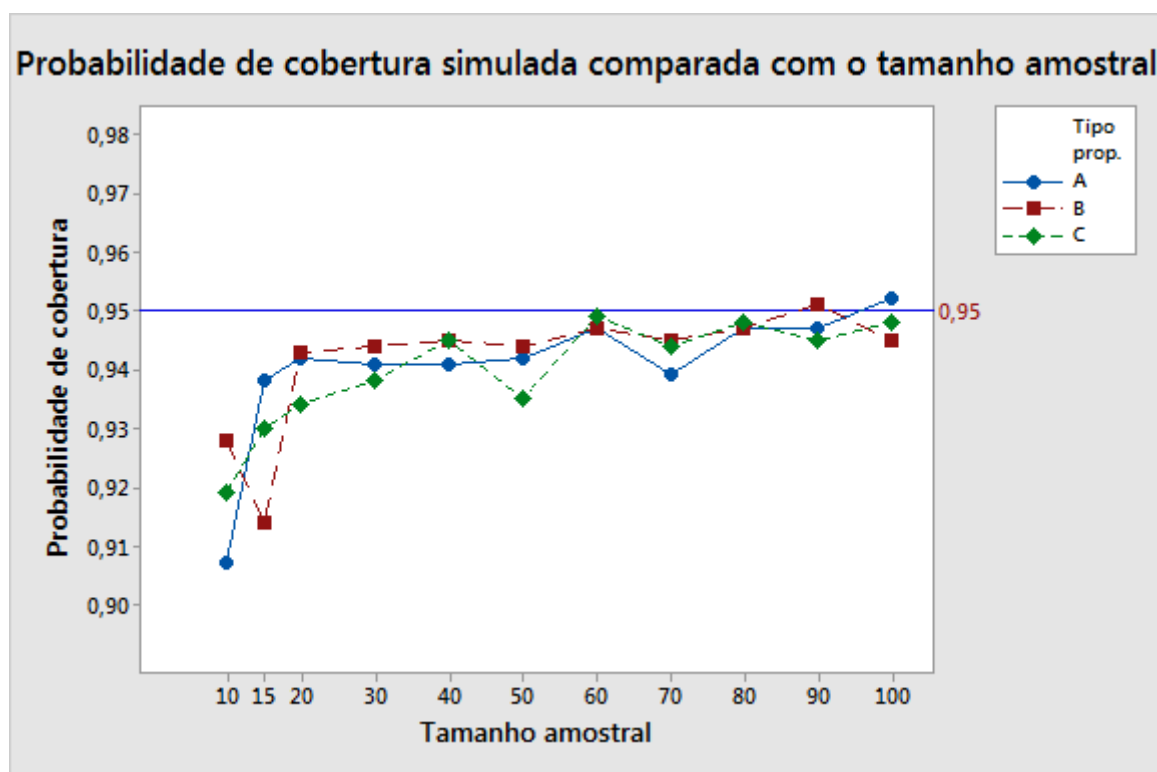


Figura 1 Probabilidades de cobertura simulada representada graficamente contra o tamanho amostral para cada categoria de populações Bernoulli.

Os resultados nas Tabelas 1-11 e na Figura 1 mostram que as amostras geradas de populações Bernoulli na categoria B (quando ambas as proporções estão próximas de 0,5), geralmente produzem probabilidades de cobertura simulada que são mais estáveis e perto da cobertura alvo de 0,95. Nesta categoria, os números esperados de sucessos e falhas em ambas as populações são maiores do que em outras categorias, mesmo quando as amostras são pequenas.

Por outro lado, para as amostras geradas dos pares de populações Bernoulli, na categoria A (quando ambas as proporções estão próximas de 1,0) ou na categoria C (quando uma proporção está perto de 1,0 e a outra próximo de 0), as probabilidades de cobertura simulada estão fora do alvo nas amostras menores, exceto quando ambos o número esperado de sucessos (np) ou o número esperado de falhas (n(1-p)) é grande o bastante.

Por exemplo, considere as amostras geradas das populações Bernoulli na categoria A em que $n = 15$. Os números esperados de sucessos são 12,0 e 13,5 e os números esperados de falhas são 3,0 e 1,5 para cada população, respectivamente. Apesar de o número esperado de falhas ser menor do que 5 para ambas as populações, a probabilidade de cobertura simuladas é de cerca de 0,94. Resultados como esses nos levam a criar a Regra 2, que requer somente que *ambos* o número esperado de sucessos *ou* o número esperado de falhas seja maior do que ou igual a 5 para cada amostra.

Para avaliar mais completamente quão eficazmente a Regra 1 e a Regra 2 pode avaliar a aproximação para o intervalo de confiança, representamos graficamente o percentual de amostras que satisfazem a Regra 1 e o percentual de amostras que satisfazem a Regra 2 contra as probabilidade da cobertura simulada nos experimentos. Os gráficos estão exibidos na Figura 2.

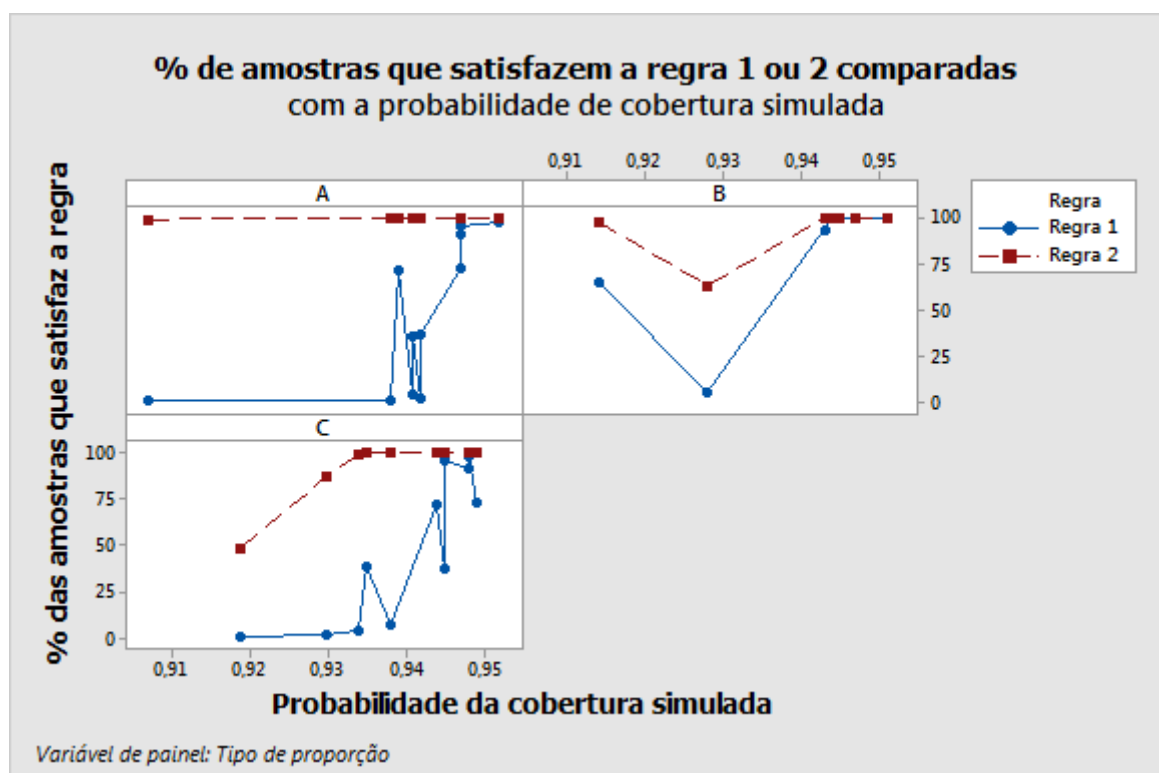


Figura 2 O percentual de amostras que satisfazem a Regra 1 e a Regra 2 representados graficamente contra a probabilidade de cobertura simulada, para cada categoria de populações de Bernoulli.

Os gráficos mostram como as probabilidades de cobertura simulada se aproximam da cobertura alvo de 0,95, o percentual das amostras que atendem os requisitos de cada regra geralmente se aproxima de 100%. Para amostras geradas de populações Bernoulli nas categorias A e C, a Regra 1 é rigorosa quando as amostras são pequenas, conforme evidenciado pelo percentual extremamente baixo de amostras que satisfaziam a regra, ainda que as probabilidades de cobertura simulada estejam próximas do alvo. Por exemplo, quando $n = 20$ e as amostras são geradas das populações Bernoulli na categoria A, a probabilidade de cobertura simulada é 0,942 (consulte a Tabela 3). Contudo, a proporção

das amostras que satisfazem a regra é praticamente 0 (0,015) (consulte a Figura 2). Portanto, nesses casos, a regra pode ser muito conservadora.

A Regra 2, por outro lado, é menos rigorosa para pequenas amostras geradas das populações Bernoulli na categoria A. Por exemplo, conforme mostrado na Tabela 1, quando $n = 10$ e as amostras são geradas de populações Bernoulli na categoria A, a probabilidade de cobertura simulada é de 0,907 e 99,1% das amostras satisfazem a regra.

Na conclusão, a Regra 1 tende a ser excessivamente conservadora quando as amostras são pequenas. A Regra 2 é menos conservadora e poder de preferida quando os tamanhos amostrais são pequenos. Entretanto, a Regra 1 é bem conhecida e bem aceita. Apesar de a Regra 2 mostrar potencial promissor, em alguns casos ela também pode ser muito liberal, conforme mostrado anteriormente. Uma possibilidade é combinar as duas regras para obter vantagem dos pontos fortes de cada regra, entretanto, essa abordagem requer investigação adicional antes que ela possa ser aplicada.

Apêndice E: Comparação do poder real versus poder teórico

Simulação E1: Estimando o poder real usando o teste exato de Fisher

Planejamos uma simulação para comparar os níveis de poder real estimado (referidos como níveis de poder simulado) do teste exato de Fisher para os níveis de poder teórico com base na função de poder do teste de aproximação normal (referido como níveis de poder aproximado). Em cada experimento, geramos 10.000 pares de amostras de pares de populações Bernoulli. Para cada par de amostras, as proporções foram escolhidas de forma que a diferença entre as proporções foi $p_1 - p_2 = -0,20$.

- **Proporções A: ambas p_1 e p_2 estão perto de 1,0 (ou próximas de 0).** Para representar este par de populações Bernoulli na simulação, usamos $p_1 = 0,70$ e $p_2 = 0,90$.
- **Proporções B: p_1 e p_2 estão próximos de 0,5.** Para representar este par de populações Bernoulli na simulação, usamos $p_1 = 0,40$ e $p_2 = 0,60$.
- **Proporções C: ambas p_1 e p_2 estão perto de 1,0.** Para representar este par de populações Bernoulli na simulação, usamos $p_1 = 0,55$ e $p_2 = 0,75$.

Fixamos os tamanhos amostrais de ambos os pares em um valor único de n , em que $n = 10, 15, 20, 30, \dots, 100$. Limitamos o estudo a experimentos balanceados ($n_1 = n_2 = n$) porque geralmente assume-se que as duas amostras têm o mesmo tamanho. Calculamos um tamanho amostral comum necessário para detectar uma diferença praticamente importante com um determinado poder.

Para estimar o poder real do teste exato de Fisher com base nos resultados de cada simulação, calculamos a fração dos 10.000 pares de amostras para os quais o teste bilateral foi significativo no nível alvo de significância, $\alpha = 0,05$. Depois, calculamos os níveis de poder teóricos correspondentes com base no teste de aproximação normal para comparação. Os resultados são apresentados na Tabela 12 abaixo.

Tabela 12 Níveis de poder simulados do teste exato de Fisher comparados com os níveis de poder aproximado das três categorias de populações Bernoulli. O nível alvo de significância é $\alpha = 0,05$.

<i>n</i>	Proporções A		Proporções B		Proporções C	
	<i>p</i> ₁ = 0,70 <i>p</i> ₂ = 0,90		<i>p</i> ₁ = 0,40 <i>p</i> ₂ = 0,60		<i>p</i> ₁ = 0,55 <i>p</i> ₂ = 0,75	
	Poder Simulado	Poder Aprox.	Poder Simulado	Poder Aprox.	Poder Simulado	Poder Aprox.
10	0,063	0,193	0,056	0,14	0,056	0,149
15	0,151	0,271	0,097	0,19	0,101	0,204
20	0,244	0,348	0,146	0,24	0,183	0,259
30	0,37	0,49	0,256	0,338	0,272	0,366
40	0,534	0,612	0,371	0,431	0,381	0,466
50	0,641	0,711	0,477	0,516	0,491	0,556
60	0,726	0,789	0,536	0,593	0,56	0,635
70	0,814	0,849	0,61	0,661	0,649	0,703
80	0,87	0,893	0,66	0,72	0,716	0,76
90	0,907	0,925	0,716	0,77	0,772	0,808
100	0,939	0,948	0,792	0,812	0,812	0,848

Os resultados na Tabela 12 mostram que o poder aproximado tende a ser maior do que o poder simulado para todas as três categorias de populações Bernoulli (A, B e C). Por exemplo, para as proporções na categoria A, o tamanho amostral real necessário para detectar uma diferença absoluta de -0,20 com um nível de poder aproximado de 0,91 é de cerca de 90. Em contraste, a estimativa do tamanho amostral correspondente com base na função do poder teórico aproximado é de cerca de 85. Portanto, a estimativa do tamanho amostral com base na função de poder aproximado é geralmente, ligeiramente menor do que o tamanho amostral real para alcançar um determinado nível de poder.

Você pode ver essa relação mais claramente quando os resultados são exibidos como curvas de poder, conforme mostrado na Figura 3 a seguir.

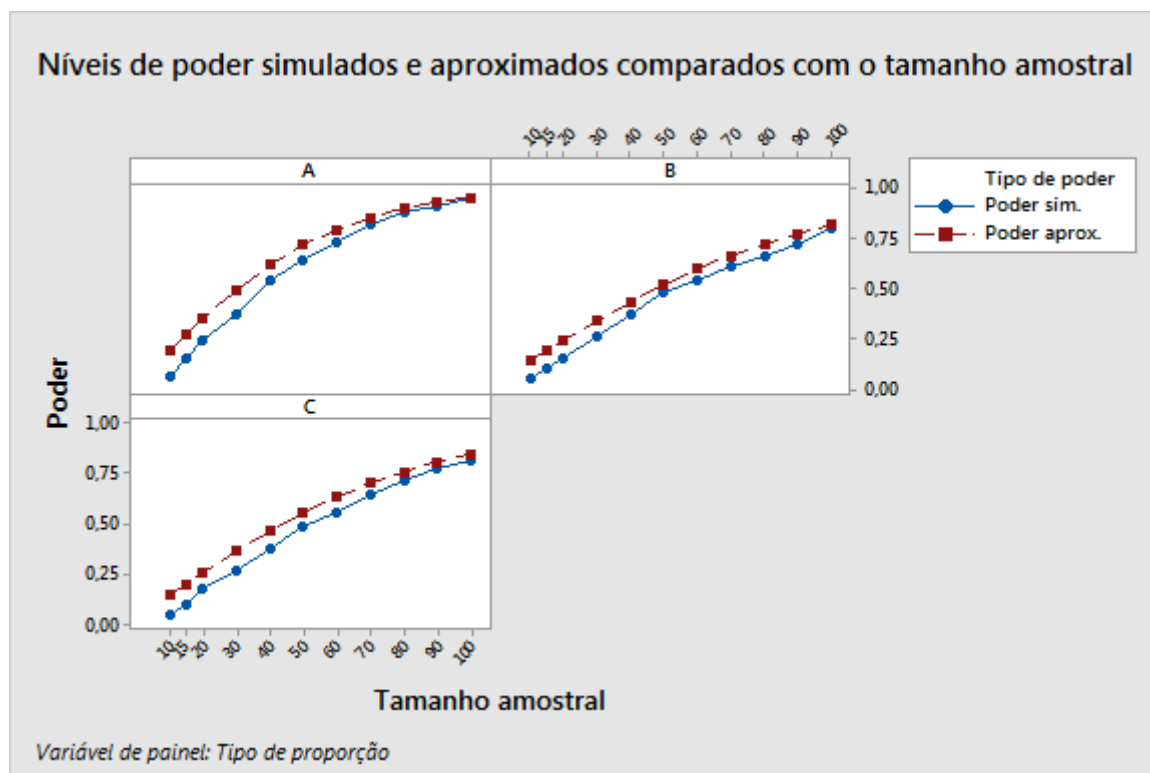


Figura 3 Gráficos de níveis de poder simulado e aproximado do teste bilateral para comparação de duas proporções. Os níveis de poder são representados graficamente contra o tamanho amostral em painéis separados para cada categoria de populações Bernoulli.

Observe que apesar de as curvas de poder simulado serem menores do que as curvas de poder aproximado para todas as três categorias de populações Bernoulli (A, B e C), o tamanho da diferença entre as curvas depende das proporções reais das populações Bernoulli das quais as amostras são traçadas. Por exemplo, quando as duas proporções estão perto de 0,5 (categoria B), os dois níveis de poder estão, geralmente, perto. Contudo, a disparidade entre as duas curvas de poder é mais notável em pequenas amostras para as proporções associadas às categorias de população A e C.

Esses resultados mostram, em geral, a função de poder teórico do teste de aproximação normal e a função de poder simulado do teste exato de Fisher são praticamente iguais. Portanto, o Assistente usa a função de poder teórico do teste de aproximação normal para estimar os tamanhos amostrais, antes de realizar o teste exato de Fisher. Portanto, os tamanhos amostrais, calculados usando-se a função de poder aproximado, podem ser um pouco menores do que os tamanhos amostrais reais, para alcançar um determinado poder para detectar uma diferença entre as duas proporções (% de defeituosos).

© 2020 Minitab, LLC. All rights reserved. Minitab®, Minitab Workspace™, Companion by Minitab®, Salford Predictive Modeler®, SPM®, and the Minitab® logo are all registered trademarks of Minitab, LLC, in the United States and other countries. Additional trademarks of Minitab, LLC can be found at www.minitab.com. All other marks referenced remain the property of their respective owners.