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.

Testes de qui-quadrado

Visão geral

Na prática, os profissionais de qualidade, por vezes, precisam coletar dados categóricos para avaliar um processo quando não é possível ou conveniente coletar dados contínuos. Por exemplo, um produto pode ser classificado em duas categorias como defeituoso/não defeituoso ou em mais de duas categorias, como excelente, bom, regular e ruim. Outro exemplo é um departamento financeiro que controla o número de dias de atraso das faturas em categorias: 15 dias ou menos, 16 a 30 dias, 31 a 45 dias ou 45 dias ou mais. Como resultado, a variável de interesse é o número de itens que se enquadram em cada categoria.

Devido à sua versatilidade, os testes qui-quadrado são usados para muitas aplicações que envolvem dados categóricos. No Assistente, usamos testes qui-quadrado para:

- Teste de qualidade de ajuste para uma distribuição multinomial

  É possível usar este teste para determinar se os dados seguem a mesma distribuição que seguiam no passado. A distribuição é definida como uma distribuição multinomial com um conjunto de porcentagens históricas, ou alvo, que definem a percentagem de itens que se enquadram em cada categoria da resposta. O qui-quadrado testa conjuntamente se qualquer percentual difere significativamente do seu respectivo percentual histórico ou alvo.

- Teste da igualdade entre % de defeituosos para mais de 2 grupos

  É possível usar este teste para determinar se existe uma diferença entre os percentuais de defeituosos dos diferentes grupos. Os grupos diferem por uma característica de interesse, como um produto produzido por diferentes operadores, por fábricas diferentes ou em momentos diferentes. O qui-quadrado testa conjuntamente se algum percentual de defeituosos difere significativamente de qualquer outro percentual de defeituosos.

- Teste da associação entre duas variáveis categóricas

  É possível usar este teste para determinar se uma variável resposta categórica (Y) está relacionada ou associada a outra variável preditora categórica (X). O qui-quadrado testa conjuntamente se existe uma associação entre a variável resposta e uma variável preditora. No Assistente, é possível realizar um Teste Qui-Quadrado de Associação com uma variável preditora (X) que contém dois ou mais valores diferentes (duas ou mais amostras).

Para obter mais detalhes sobre a estatística de teste qui-quadrado, consulte o Anexo A.

Para os métodos que envolvem o teste de hipóteses, é uma boa prática garantir que as premissas para o teste sejam satisfeitas, que o teste tenha poder suficiente e que quaisquer aproximações usadas para analisar os dados produzam resultados válidos. Para os testes qui-quadrado, os pressupostos são inerentes à coleta de dados e nós não tratamos deles nas verificações de dados.

Concentramos nossa atenção sobre o poder e a validade dos métodos de aproximação. O Assistente usa esses métodos de aproximação para realizar as seguintes verificações em seus dados e reporta os resultados do Relatório do Cartão:

- Tamanho amostral
- Validade do teste
- Validade dos intervalos

Neste trabalho, investigamos como essas verificações de dados estão relacionadas a testes de qui-quadrado na prática, e descrevemos como foram estabelecidas as diretrizes para as verificações de dados no Assistente.

Verificações dos dados

Tamanho amostral

Normalmente, o principal objetivo para a realização de um testes estatísticos de hipóteses é reunir evidências para rejeitar a hipótese nula de "nenhuma diferença". Se as amostras forem pequenas demais, talvez o poder do teste não seja suficiente para detectar uma diferença entre os percentuais de defeituosos que realmente exista, o que resulta um erro tipo II. Assim, é fundamental garantir que os tamanhos de amostra sejam suficientemente grandes para detectar diferenças práticas importantes com alta probabilidade.

A verificação dos dados do tamanho da amostra baseia-se na poder do teste. Este cálculo requer que o usuário especifique uma diferença significativa entre um parâmetro de população real e o valor nulo hipotético. Como era muito difícil determinar e expressar essa diferença prática para o Teste de Ajuste Qui-Quadrado e o Teste Qui-Quadrado para Associação, o Assistente só verifica o tamanho da amostra para o teste de % Defeituosos Qui-Quadrado com mais de duas amostras.

Objetivo

Se os dados não fornecem evidências suficientes contra a hipótese nula, queremos determinar se os tamanhos de amostra são grandes o suficiente para o teste detectar diferenças práticas de interesse com alta probabilidade. Embora o objetivo do planejamento do tamanho da amostra seja garantir que os tamanhos de amostra são grandes o suficiente para detectar diferenças importantes com grande probabilidade, as amostras não devem ser tão grandes de forma que diferenças insignificantes se tornem estatisticamente significativas com alta probabilidade.

Método

A análise do tamanho da amostra e do poder está baseada nas fórmulas mostradas no Anexo B.

Resultados

Quando os dados não fornecem evidências suficientes contra a hipótese nula e você não especificou uma diferença prática, o Assistente calcula as diferenças práticas que podem ser detectadas com 80% e 90% de probabilidade com base nos tamanhos de amostra. Além disso, se o usuário fornecer uma diferença prática específica de interesse, o Assistente calcula os tamanhos das amostras que produzem 80% e 90% de chance de detectar essa diferença.

Ao verificar o poder e tamanho da amostra, o Cartão de Relatório do Assistente para o teste Qui-Quadrado de % Defeituosos para mais de duas amostras exibe os seguintes indicadores de status:

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 os % de defeituosos, mas a amostra é grande o suficiente para fornecer pelo menos 90% de chance de detectar a determinada diferença.
	O poder pode ser suficiente. O teste não encontrou diferença entre os % de defeituosos, mas a amostra é grande o suficiente para oferecer uma chance de 80% a 90% de detectar determinada diferença. O tamanho da amostra necessária para atingir 90% de poder é informado.
	Talvez o poder não seja suficiente. O teste não encontrou diferença entre os % de defeituosos, e a amostra é grande o suficiente para proporcionar uma chance de 60% a 80% de detectar determinada diferença. Os tamanhos das amostras necessárias para atingir 80% e 90% de poder são relatados.
	O poder não é suficiente (< 60%). O teste não encontrou diferença entre os % de defeituosos. Os tamanhos das amostras necessárias para atingir 80% e 90% de poder são relatados.
	O teste não encontrou diferença entre os % de defeituosos. Você não especificou uma diferença prática entre os % de defeituosos a serem detectados; portanto, o relatório indica as diferenças que poderiam ser detectadas com 80% e 90% de chance, com base em seus tamanhos de amostra e alfa.

Validade do teste

A estatística de testeχ^2 apenas segue aproximadamente uma distribuição qui-quadrado. A aproximação melhora com amostras maiores. Nesta seção, avaliamos a aproximação usada para determinar o tamanho mínimo da amostra necessária para obter resultados precisos.

A aproximação qui-quadrado para a estatística de teste é avaliada através da análise do impacto da baixa contagem esperada nas células sobre a taxa de erro tipo I (alfa). Ao utilizar o erro tipo I para avaliar a validade do teste, nós desenvolvemos uma regra para garantir que:

- A probabilidade de rejeição da hipótese nula quando ela é verdadeira é pequena e próxima da taxa de erro tipo I desejada.
- A cauda da distribuição nula pode ser razoavelmente aproximada, o que é importante para calcular com precisão o valor-p.

Utilizando uma abordagem padrão, definimos uma contagem baixa esperada de célula como uma célula que tem uma contagem esperada inferior ou igual a 5.

Foram desenvolvidos dois modelos para definir as proporções da hipótese nula: o modelo de proporções perturbadas e o modelo de proporção igual. Para obter mais detalhes, consulte o Anexo C. Ambos os modelos são usados nas simulações mencionadas mais adiante neste documento. Os modelos são usados para cada um dos testes qui-quadrado, com uma exceção: o modelo de proporções perturbadas não se aplicam ao teste Qui-Quadrado de % Defeituosos para mais de duas amostras.

A Validade do Teste de verificação de dados se aplica a todos os testes de qui-quadrado no Assistente. Cada verificação de dados é descrita abaixo.

Teste Qui-Quadrado de Qualidade do Ajuste

Objetivo

Avaliamos a aproximação qui-quadrado para a estatística de teste investigando o impacto da magnitude e a frequência das contagens pequenas esperadas na taxa de erro tipo I.

Método

Amostras de tamanho n foram retiradas de uma distribuição multinomial com as proporções descritas nas proporções perturbadas ou modelos de proporções iguais (consulte Anexo C). Para cada condição, foram realizados 10.000 testes de qualidade de ajuste de qui-quadrado com um nível de significância alvo de 0,05. Para cada teste, calculamos o erro real do tipo I como $\frac{\textit{Número de testes rejeitados}}{\textit{Número de réplicas}\ (10000)}$. Nós definimos o intervalo para as taxas de erro tipo I aceitáveis [0,03-0,07] e registramos o tamanho mínimo da amostra com uma taxa erro tipo I nesse intervalo.

Resultados

Os resultados das simulações mostraram que a contagem de células alvo menor do que 1,25 pode conduzir a valores de p incorretos quando a porcentagem de contagens de células alvo pequenas for menor ou igual a 50%. Além disso, a contagem de célula alvo menor do que 2,5 pode conduzir a valores de p incorretos quando a porcentagem de contagens de células alvo pequenas for superior a 50%. Consulte o Anexo D para obter mais detalhes.

Quando verificar a validade da Qualidade de Ajuste do qui-quadrado, o Cartão de Relatório do Assistente exibe os indicadores a seguir:

Status	Condição
	A contagem mínima de células alvo é maior ou igual a 1,25 quando a porcentagem da contagem de células alvo pequenas for menor ou igual a 50% OU A contagem mínima de células alvo é maior ou igual a 2,5 quando a porcentagem da contagem de células alvo pequenas for maior do que 50%. Sua amostra é grande o suficiente para obter contagens alvo suficientes. O valor-p para o teste deve ser preciso.
	Se as condições acima não forem sustentadas.

Teste Qui-Quadrado para Associação

Objetivo

Avaliamos a aproximação qui-quadrado para a estatística de teste investigando o impacto da magnitude e a frequência das contagens pequenas esperadas na taxa de erro tipo I.

Método

Os tamanhos de amostra $n_i$são provenientes de uma distribuição multinomial com as proporções definidas pelos modelos de proporções perturbadas ou proporções iguais (consulte o Anexo C). Para simplificar, nós escolhemos $n_i = n \; \forall i$. Para cada condição, foram realizados 10.000 testes de qui-quadrado para associação com um nível de significância alvo de 0,05. Para cada teste, calculamos a taxa de erro tipo I como real. Nós definimos o intervalo para as taxas de erro tipo I aceitáveis [0,03-0,07] e registramos o tamanho mínimo da amostra com uma taxa erro tipo I nesse intervalo.

Resultados

Descobrimos que a contagem de células mínima esperado depende do número de valores de X e da porcentagem de contagens de células pequenas esperadas.

- Para o modelo de proporções perturbadas, quando a percentagem de contagens de células pequenas esperadas for menor ou igual a 50% de perturbação, as contagens de células mínimas esperadas são ≤ 2 e ≤ 1 para o número de valores de X, são iguais a (2 ou 3) e (4, 5 ou 6), respectivamente. Além disso, quando a porcentagem de contagens de células pequenas esperadas for > 50%, as contagens de células mínimas esperadas serão ≤ 3 e ≤ 1,5 para o número de valores de X igual a (2 ou 3) e (4, 5, ou 6), respectivamente.
- Para o modelo de proporções iguais, a contagem de células mínima esperada é ≤ 2, quando o número de valores de X for igual a (2 ou 3) e a contagem de células mínima esperada é ≤ 1,5 quando o número de valores de X é igual a (4, 5 ou 6).

Para obter mais detalhes, consulte o Anexo E.

Quando verificar a validade da teste qui-quadrado para associação, o Cartão de Relatório do Assistente exibe os indicadores a seguir:

Status	Número de valores da variável X	Condição
	2 ou 3	A contagem mínima de células esperada é maior ou igual a 2 quando a porcentagem da contagem de células alvo pequenas esperada (menor ou igual a 5) for menor ou igual a 50% A contagem mínima de células esperada é maior ou igual a 3 quando a porcentagem da contagem de células alvo pequenas esperada (menor ou igual a 5) for maior do que 50%

Status	Número de valores da variável X	Condição
	4, 5 ou 6	A contagem mínima de células esperada é maior ou igual a 1 quando a porcentagem da contagem de células alvo pequenas esperada (menor ou igual a 5) for menor ou igual a 50% A contagem mínima de células esperada é maior ou igual a 2 (com arredondamento conveniente de 1,5 para 2) quando a porcentagem da contagem de células alvo pequenas esperada (menor ou igual a 5) for maior do que 50%
	Todos os casos	Se as condições acima não forem sustentadas.

Teste qui-quadrado para % de defeituosos com mais de duas amostras

Objetivo

Avaliamos a aproximação qui-quadrado para a estatística de teste investigando o impacto da magnitude e a frequência das contagens pequenas esperadas na taxa de erro tipo I.

Método

Definimos o modelos p = $p_i \ = \ p_j \ \forall i, j$, em que p = 0,001, 0,005, 0,01, 0,025 e 0,25. Os tamanho de amostra n_i foram extraídos de uma distribuição binomial com os valores de p_i descritos acima. Para simplificar, nós escolhemos $n_i = n \ \forall i$. Para cada condição, foram realizados 10.000 testes qui-quadrado para o % de defeituosos com um nível de significância alvo de 0,05. Para cada teste, calculamos o erro real do tipo I como $\frac{Número\ de\ testes\ rejeitados}{Número\ de\ réplicas\ (10000)}$. Nós definimos o intervalo para as taxas de erro tipo I aceitáveis [0,03 - 0,07] e registramos o tamanho mínimo da amostra com uma taxa de erro tipo I nesse intervalo.

Resultados

Quando existem 3 a 6 valores de X, um número mínimo esperado de defeituosos e de não defeituosos maior ou iguais a 1,5 produz uma taxa de erro de Tipo I para o teste no intervalo [0,03, 0,07]. Quando existem 7 a 12 valores de X, um número mínimo esperado de defeituosos e não defeituosos superior ou igual a 1 origina uma taxa de erro de Tipo I para o teste no intervalo [0,03, 0,07].

Para obter mais detalhes, consulte o Anexo F.

Quando da verificação da validade do teste qui-quadrado para % de defeituosos para mais de duas amostras , o Cartão de Relatório do Assistente exibe os seguintes indicadores de status:

Status	Número de valores de X	Condição
	3 a 6	O número mínimo esperado de defeituosos e de não defeituosos maior ou iguais a 1,5.
	7 a 12	O número mínimo esperado de defeituosos e de não defeituosos maior ou iguais a 1.
	Todos os casos	Se as condições acima não forem sustentadas.

Validade dos intervalos

Os intervalos de comparação o teste qui-quadrado para % de defeituosos para mais de duas amostras e teste de Qualidade de Ajuste do Qui-quadrado são baseados na aproximação normal. Além disso, os intervalos de confiança individuais no teste de Qualidade de Ajuste do Qui-quadrado são baseados na aproximação normal. Nesta seção, nós avaliamos a validade da aproximação normal. De acordo com a regra geral encontrada na maioria dos livros didáticos de estatística, o intervalo de confiança aproximado é preciso se as contagens observadas forem pelo menos 5.

A validade da verificação dos intervalos de dados aplica-se a teste qui-quadrado para % de defeituosos defeituoso com mais de duas amostras e teste de Qualidade de Ajuste do Qui-quadrado.

Teste qui-quadrado para % de defeituosos com mais de duas amostras

Objetivo

Queríamos avaliar a regra geral para o número mínimo de defeituosos e não defeituosos observados em cada amostra para garantir que os intervalos de confiança aproximados sejam precisos.

Método

Primeiramente, definimos os intervalos que são utilizados no gráfico de comparação. Os pontos finais dos intervalos são definidos de modo que, com uma taxa de erro global de aproximadamente $\propto$, qualquer intervalo que não consiga se sobrepor indique a população % de defeituosos que são diferentes. Consulte o Anexo G para conhecer as fórmulas utilizadas.

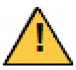
Os intervalos de comparação são baseadas em intervalos de confiança de comparação pareada. Para obter mais detalhes, consulte a seção de intervalos de comparação no White

Paper do Assistente para ANOVA de um fator. Usamos um intervalo de confiança aproximação normal para cada par (pi – pj) e, em seguida, usamos um procedimento de comparações múltiplas de Bonferroni para controlar a taxa de erro geral na mesma direção do experimento. Portanto, só precisamos avaliar a validade de um dos intervalos no procedimento de comparação pareada para entender o efeito da aproximação normal sobre os intervalos de comparação.

Resultados

Para avaliar a validade da aproximação normal, só precisamos examinar como a aproximação afeta um intervalo para a diferença entre os % de defeituosos. Portanto, podemos simplesmente usar a regra geral desenvolvida para casos de % de defeituosos com duas amostras. Para obter mais detalhes, consulte a seção de métodos de teste para % de defeituosos com duas amostras no White Paper do Assistente para o testes de % de defeituosos com duas amostras. Os resultados da simulação em teste de % de defeituosos com duas amostras indicam que a precisão do intervalo de confiança aproximado para a diferença entre a os % de defeituosos geralmente é confiável 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 pelo menos 5.

Quando da verificação da validade dos intervalos para o teste qui-quadrado para % de defeituosos para mais de duas amostras , o Cartão de Relatório do Assistente exibe os seguintes indicadores de status:

Status	Condição
	Todas as amostras têm pelo menos 5 defeituosos e 5 não defeituosos. Os intervalos de comparação devem ser precisos.
	Se a condição acima não vigorar.

Teste Qui-Quadrado de Qualidade do Ajuste

Objetivo

Queríamos avaliar a regra geral para o número mínimo de defeituosos e não defeituosos observados em cada amostra para garantir que os intervalos de confiança aproximados sejam precisos.

Método

Teste de Qualidade de Ajuste do Qui-quadrado do Assistente inclui a comparação e intervalos de confiança individuais. Nós utilizamos os intervalos normais de aproximação padrão para proporções e para corrigir vários intervalos usando a correção de Bonferroni (Goodman, 1965). Assim, os intervalos de Bonferroni simultâneos são calculados como a seguir:

$$p_{iInferior} = p_i - Z_{\alpha/2k}\sqrt{\frac{\mathrm{p_i}(1 - \mathrm{p_i})}{\mathrm{N}}}$$

$$p_{iSuperior} = p_i + Z_{\alpha/2k}\sqrt{\frac{\mathrm{p_i}(1 - \mathrm{p_i})}{\mathrm{N}}}$$

Os limites dos intervalos são definidos de modo que, com uma taxa de erro global de aproximadamente ∝, qualquer intervalo que não contenha o valor de proporção alvo indique que a proporção real é diferente da sua proporção alvo correspondente. Os intervalos individuais utilizam a mesma forma que os intervalos de Bonferroni, mas não são corrigidos para os intervalos múltiplos usando $Z_{\alpha/2}$.

Resultados

Ambas as abordagens descritas acima seguem uma metodologia semelhante à que foi definida no teste de % de defeituosos com 2 amostras do Assistente. Portanto, podemos usar regras semelhantes para a validade da aproximação normal que tenham sido desenvolvidas para esse teste. Para obter mais detalhes, consulte a seção no White Paper do Assistente para o teste de % de defeituosos com 2 amostras. Naquele artigo, concluímos que os intervalos de comparação e os intervalos de confiança individuais podem não ser precisos quando as contagens amostrais forem inferiores a 5.

Quando verificar a validade dos intervalos do teste de Qualidade de Ajuste do qui-quadrado, o Cartão de Relatório do Assistente exibe os indicadores a seguir:

Status	Condição
	Todas as contagens de amostra são de pelo menos 5. Os intervalos devem ser precisos.
	Existem contagens de amostra menores do que 5.

Referências

Agresti, A. (1996). An introduction to categorical data analysis. New York, NY: Wiley.

Read, T. & Cressie, N. (1988). Goodness-of-fit statistics for discrete multivariate data. New York, NY: Springer-Verlag.

Fienberg, S. (1980). The analysis of cross-classified categorical data. Cambridge, MA: MIT Press.

Goodman, L. (1965). On simultaneous confidence intervals for multinomial proportions. *Technometrics, 7*, 247-254.

Anexo A: Estatística de teste qui-quadrado

O Assistente usa uma estatística de teste do qui-quadrado da forma:

$$x^2 = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

em que

O_{ij} = contagens observadas, conforme definido na tabela abaixo:

Caso	$\boldsymbol{O_{ij}}$
Teste de qualidade de ajuste para uma distribuição multinomial.	A contagem observada para o resultado de $i^{ésimo}$ é definida como O_{i1}.
Teste da igualdade de mais de 2 % de defeituosos	O número observado de itens defeituosos e itens não defeituosos para a amostra de $i^{ésimo}$ é definido como O_{i1} e O_{i2} respectivamente.
Teste da associação entre duas variáveis categóricas	As contagens observadas para o valor de $i^{ésimo}$ da variável X e o valor de $j^{ésimo}$ da variável Y são definidas como O_{ij}.

E_{ij} = Contagem esperada conforme definido na tabela abaixo:

Caso	$\boldsymbol{E_{ij}}$
Teste de qualidade de ajuste para uma distribuição multinomial	E_{i1} = np_i $i = 1, \ldots, k$ (k = número de resultados) n = tamanho da amostra p_i = proporções históricas $\sum_i p_i = 1$
Teste da igualdade de mais de 2 % de defeituosos	E_{i1} = $n_i p$ (para defeituosos) E_{i2} = $n_i(1 - p)$ (para não defeituosos) $i = 1, \ldots, k$ (k = número de amostras) n_i = $i^{ésimo}$ tamanho da amostra p = proporção geral de defeituosos

Caso	$\boldsymbol{E_{ij}}$
Teste a associação entre duas variáveis categóricas	$E_{ij} = \frac{(n_{i.}n_{.j})}{n_{..}}$ $i = 1, \ldots, m$ (m = número de valores de X) $j = 1, \ldots, k$ (k = número de valores de Y) $n_{i.}$ = contagem total para o valor de $i^{ésimo}$ da variável X $n_{.j}$ = contagem total para o valor de $j^{ésimo}$ da variável Y $n_{..}$ = tamanho geral da amostra

Anexo B: Poder do teste qui-quadrado para % de defeituosos com mais de duas amostras

Usamos uma distribuição do Qui-quadrado não central para calcular o poder do teste que $p_i = p_j = p \; \forall i, j$. O parâmetro de não centralidade depende de n_i e de $p_i \forall i$.

em que

n_i = o tamanho de amostra para a amostra de $i^{ésimo}$

Cada p_i representa uma proporção alternativa (veja a próxima seção deste Anexo, Cálculo das Proporções Alternativas), calculada a partir da diferença de proporção = δ.

Nós calculamos o parâmetro de não centralidade da distribuição do qui-quadrado como:

$$\chi^2 = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

em que

$$O_{i1 = } n_i p_i$$

$$O_{i2 =} n_i (1 - p_i)$$

e calculamos o poder do teste como

$$\text{Prob}(X \geq x_{1-\alpha} \mid \chi^2)$$

em que

X = é uma variável aleatória de uma distribuição de qui-quadrado não central com parâmetro de centralidade χ^2.

$x_{1-\alpha}$ = fda inversa avaliada em $1 - \alpha$ para uma distribuição de qui-quadrado central.

Cálculo das Proporções Alternativas

Nós definimos as proporções alternativas como a seguir:

$$p_i = p_c + \frac{n_j}{n_i + n_j} \delta$$

$$p_j = p_c - \frac{n_i}{n_i + n_j} \delta$$

$$p_m = p_c \forall m \neq i, j$$

$$0 < \delta < 1$$

em que

$$p_c = \frac{1}{N_T} \sum_{i=1}^{k} n_i \widehat{p_i}$$

$\widehat{p_i}$ = os itens defeituosos na proporção amostral de itens defeituosos para $i^{ésimo}$.

NT = número de observações total.

n_i = tamanho da amostra para a amostra de $i^{ésimo}$.

Para algumas diferenças δ, $p_i > 1$ ou $p_j < 0$. Portanto, desenvolvemos as regras a seguir:

Se p_j < 0

$$p_i = \delta$$
$$p_j = 0$$
$$p_m = \frac{\delta}{2} \ \forall m \neq i, j$$

Se p_i > 1

$$p_i = 1$$
$$p_j = 1 - \delta$$
$$p_m = 1 - \frac{\delta}{2} \ \forall m \neq i, j$$

Usando os dois menores valores de resultados de n_i no poder mínimo e utilizando os dois maiores valores de resultados de n_i no poder máximo.

Anexo C: Modelo de proporções perturbadas e modelo de proporções iguais

Modelo de proporções perturbadas

Seguindo Read e Cressie (1988), definimos o conjunto de proporções sob a hipótese nula da seguinte forma:

Escolhemos o k - 1 próximo deδ (em que k = número de proporções de cada amostra) e definimos um conjunto de p_i pequenas na faixa de

$p_i = \frac{(1-\frac{\delta}{k-1})}{k}$ para $i = 1, \dots, r$

e o p_i restante como

$p_i = \frac{(1-\sum_{i=1}^{r} p_i)}{(k-r)}$ para $i = r + 1, \dots, k$

Os valores que usamos para δ nas simulações estão listados na Tabela 1.

Tabela 1 δ usado nas simulações com o p_i pequeno resultante

k	δ	$p_{i=1,\dots,r}$
3	1,95	0,008
4	2,95	0,004
5	3,90	0,005
6	4,90	0,003

Para cada k, nós variamos r = 1,..., k – 1 para alterar o tamanho do conjunto de p_i's. pequeno. Por exemplo, para k = 3, obtivemos os dois modelos descritos a seguir na Tabela 2.

Tabela 2 Os valores de p_i para k = 3 usando o modelo de proporções perturbadas

r	p1	p2	p3
1	0,008	0,496	0,496
2	0,008	0,008	0,984

Modelo de proporções iguais

Para se obter um modelo em que 100% das contagens de células esperadas são pequenas, utilizamos um modelo de proporção igual definido por

$$p_i = \frac{1}{k} \forall \, i$$

Utilizando este modelo, com um tamanho muito reduzido da amostra, todas as contagens de células esperadas são consideradas pequenas. Com um modelo de proporção igual, os tamanhos de amostra precisam ser muito pequenos para alcançar um número de células esperado pequenas, o que provavelmente não ocorrerá na prática.

Anexo D: Validade do teste qui-quadrado de qualidade de ajuste

Para o modelo de proporções perturbadas, foram plotados o número de células mínimas esperado necessário para alcançar uma taxa de erro de Tipo I, no intervalo [0,03, 0,07] contra o % das contagens de células pequenas esperadas, como mostrado na Figura 1.

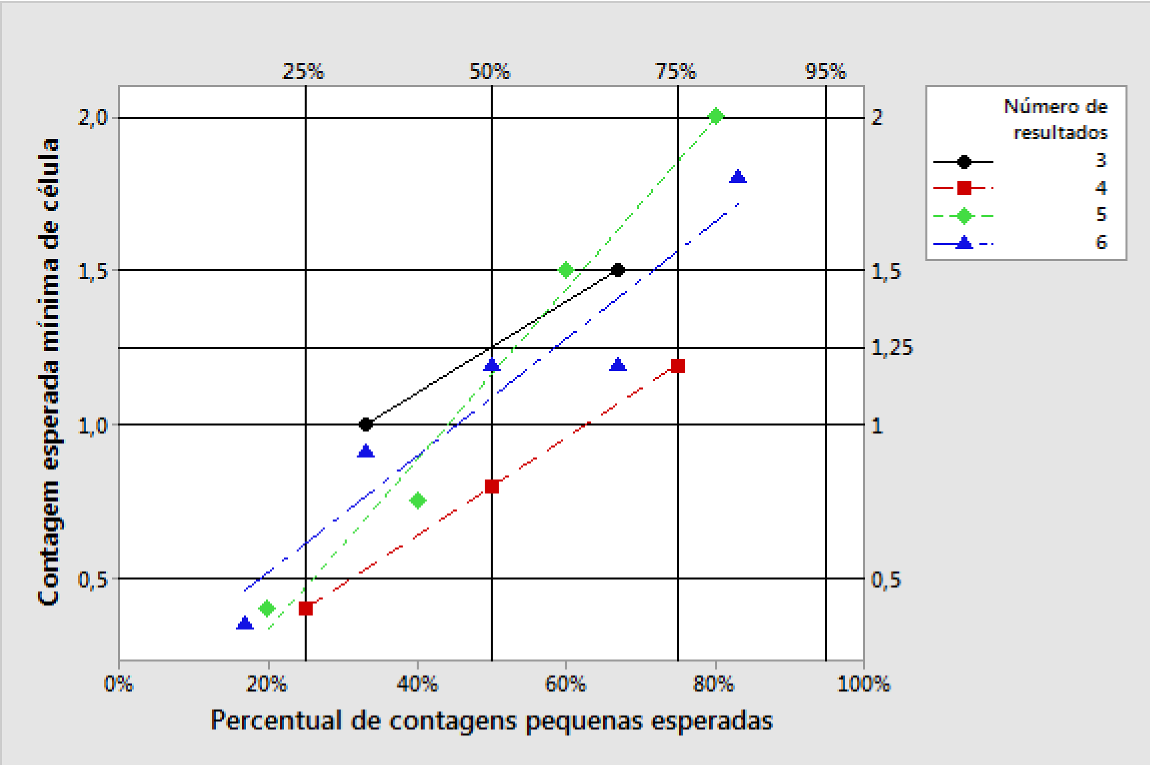


Figura 1 Contagem de células mínimas esperada necessária para atingir uma taxa de erro tipo I no intervalo [0,03, 0,07] contra a porcentagem da contagem de células pequenas esperada.

Na Figura 1, quando a percentagem de contagens de células pequenas esperadas for inferior a 50%, as contagens de células mínimas esperadas são menores ou iguais a 1,25. Todas as contagens de células mínima esperada são menores ou iguais a 2. Com base nestes resultados de simulação, as regras que usamos no Assistente do Cartão de Relatório são conservadoras.

Em seguida, foi realizada a mesma simulação utilizando o modelo de proporções iguais para definir a distribuição nula. A Tabela 4 resume os resultados da simulação utilizando um modelo de proporções iguais.

Tabela 4 Contagem de células mínimas esperada para se obter uma taxa de erro de Tipo I, no intervalo [0,03, 0,07]

k	Contagem esperada mínima de célula
3	2,5
4	1,25
5	1
6	1,4

Como indicado acima, o modelo de proporções iguais leva a casos em que 100% das contagens de células são pequenas. A Tabela 4 mostra que todas as contagens de células mínimas esperadas é menor ou igual a 2,5, o que sustenta as regras que usamos no Assistente do Cartão de relatório.

Anexo E: Validade do teste para o teste qui-quadrado para associação

Para o modelo de proporções perturbadas, foram plotados o número de células mínimas esperado necessário para alcançar uma taxa de erro de Tipo I, no intervalo [0,03, 0,07] contra o % das contagens de células pequenas esperadas para cada número de valores de X, como mostrado na Figura 2.

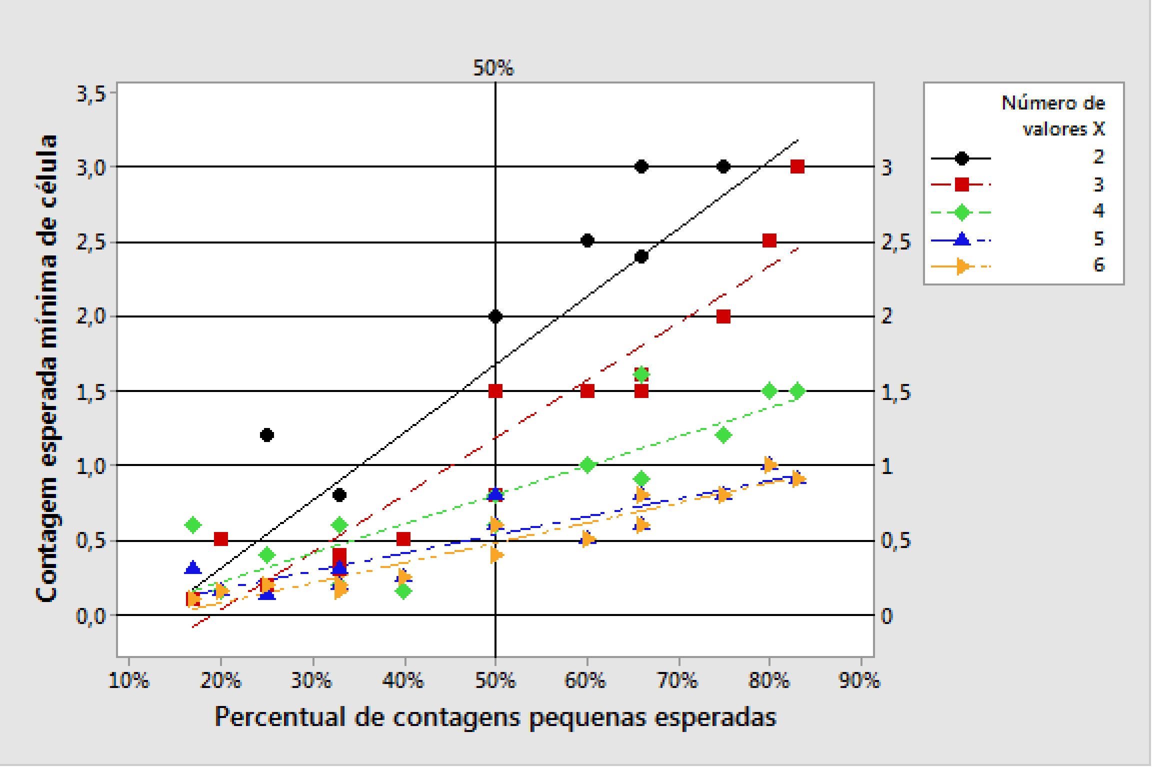


Figura 2 Contagem de células mínimas esperada necessária para atingir uma taxa de erro tipo I no intervalo [0,03, 0,07] contra a porcentagem da contagem de células pequenas esperada.

A figura 2 indica que a contagem de células mínima esperada depende do número de valores de X e da porcentagem de contagens de células pequenas esperadas.

A Figura 2 indica que, quando a porcentagem de contagens de células pequenas esperadas é ≤ 50%, as contagens de células mínimas esperadas são ≤ 2 e ≤ 1 para os números de valores de X iguais a 2 ou 3 e 4, 5, ou 6, respectivamente. Além disso, quando a porcentagem de contagens de células pequenas esperadas é > 50%, as contagens de células mínimas esperadas são ≤ 3 e ≤ 1,5 para o número de valores de X iguais a 2 ou 3 e 4, 5, ou 6, respectivamente.

Para o modelo de proporções iguais, foram plotadas a contagem de células mínimas esperadas contra o número de valores de X (m) e o número de valores de Y (k), como mostrado na Figura 3.

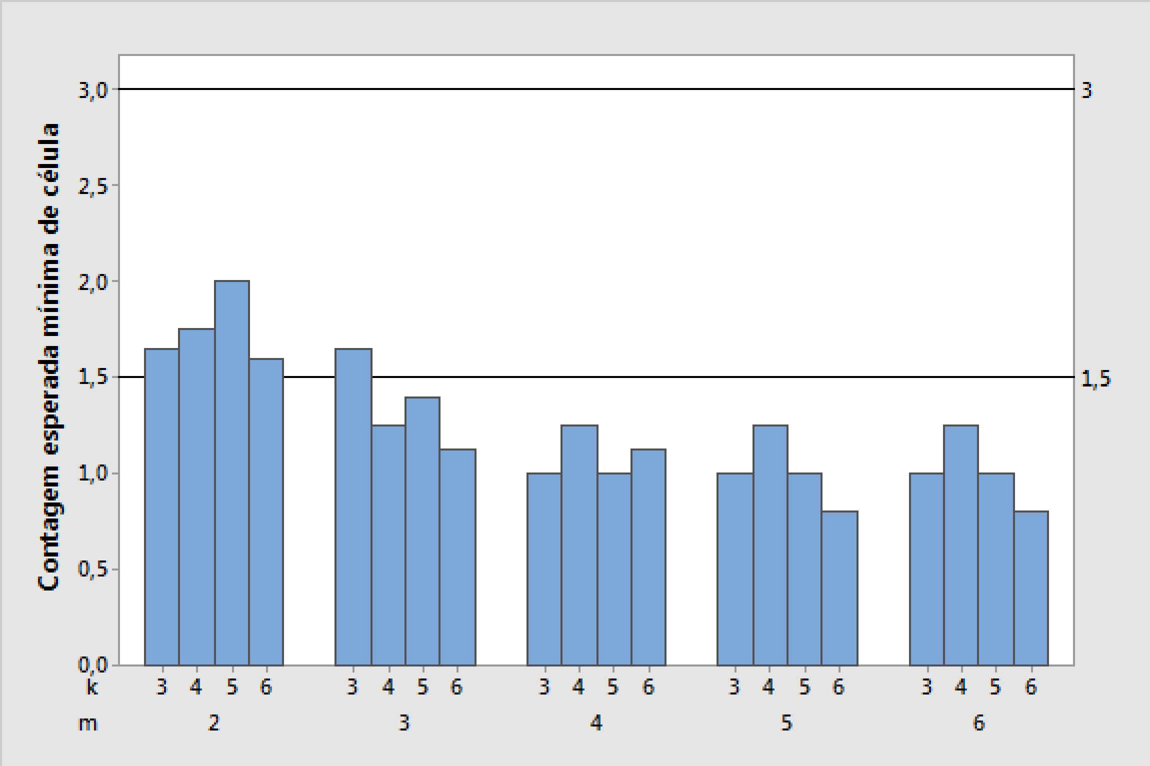


Figura 3 A contagem de células mínimas esperada necessária para atingir uma taxa de erro de Tipo I, no intervalo [0,03, 0,07] versus os valores de X (m) e os valores Y (k)

A Figura 3 indica que a contagem de células mínimas esperada é ≤ 2, quando o número de valores de X é igual a 2 ou 3 e a contagem de células mínimas esperadas é ≤ 1,5 quando o número de valores de X é igual a 4, 5 ou 6. Com base nestes resultados de simulação, as regras do Assistente do Cartão de Relatório são conservadoras.

Anexo F: Validade do teste qui-quadrado para % de defeituosos com mais de duas amostras

Para cada p e cada m = 3, 4, 5, ..., 12, plotamos a contagem esperada de células mínimas. Os resultados são apresentados nas Figuras 4 e 5.

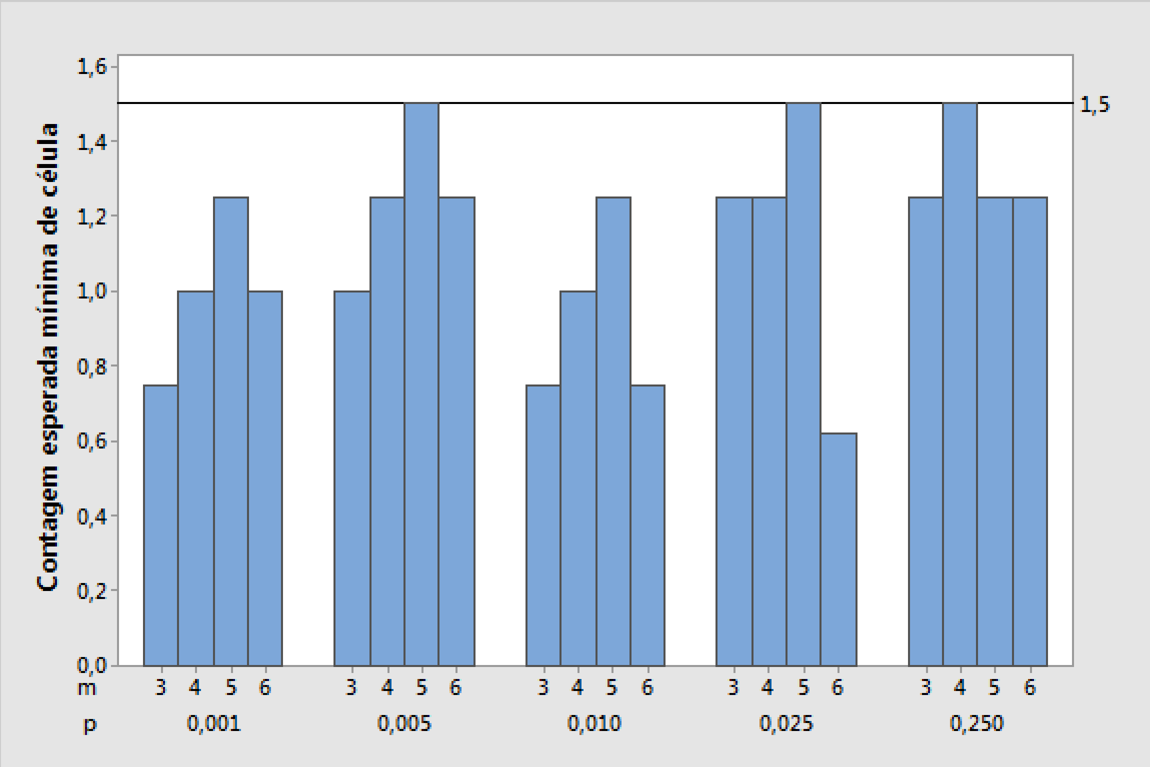


Figura 4 A contagem de células mínimas esperada necessária para atingir uma taxa de erro de Tipo I, no intervalo [0,03, 0,07] versus os número dos valores de X (m = 3 a 6)

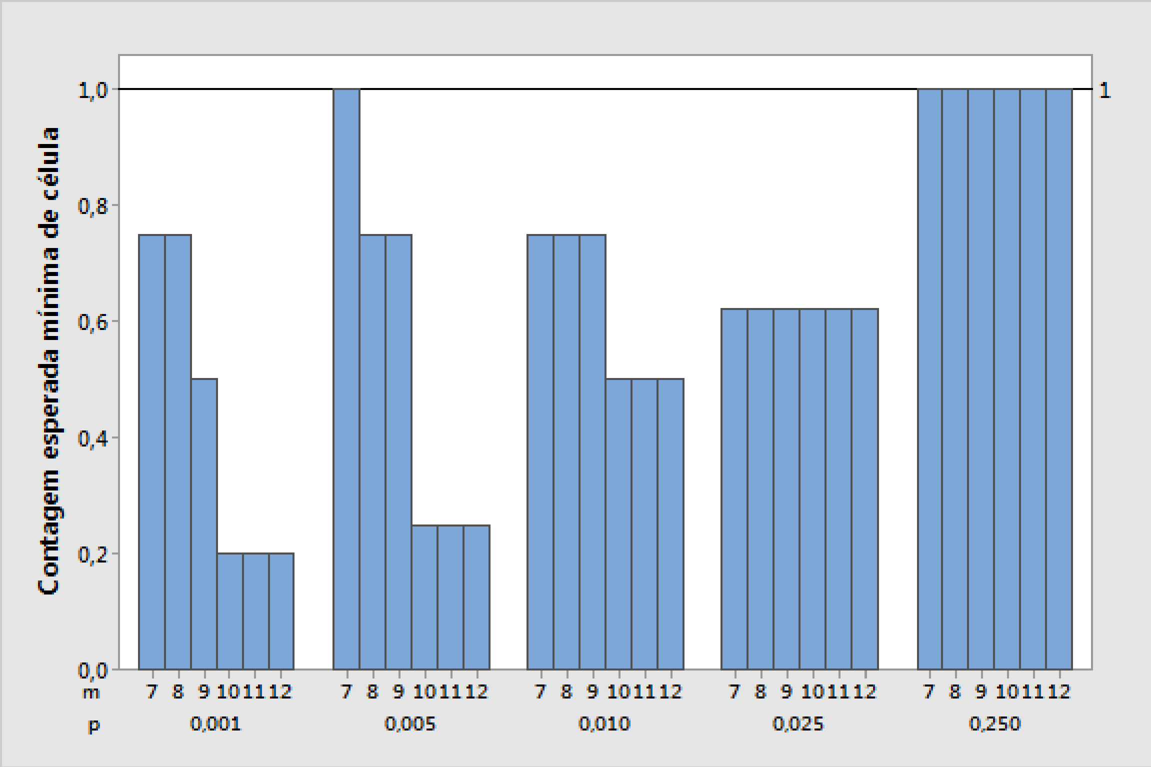


Figura 5 A contagem de células mínimas esperada necessária para atingir uma taxa de erro de Tipo I, no intervalo [0,03, 0,07] versus os número dos valores de X (m = 7 a 12)

Quando o número de valores de X é igual a 3, 4, 5 ou 6, uma contagem célula esperada maior ou igual a 1,5 produz uma taxa de erro de Tipo I para o teste no intervalo [0,03, 0,07]. Quando o número de valores de X igual a 7, 8, 9, ..., 12, uma contagem de célula esperada maior ou igual a 1 origina uma taxa de erro de Tipo I para o teste no intervalo [0,03, 0,07].

Anexo G: Comparação dos intervalos para o teste qui-quadrado para % de defeituosos com mais de duas amostras

Os limites inferior e superior para p_i são definidos da seguinte maneira:

$$p_{iInferior} = p_i - Z_{\alpha/c} X_i$$

$$p_{iSuperior} = p_i + Z_{\alpha/c} X_i$$

em que

c = número de comparações = k (k - 1) /2

em que k é o número de amostras

$Z_{\alpha/c}$ = $(1 - \frac{\alpha}{2c})$ o percentil de uma distribuição normal com média = 0 e desvio padrão = 1

$$X_i = ((k-1)\sum_{j \neq i} b_{ij} - \sum\sum_{1 \leq j < l \leq k} b_{jl}) \,/\, ((k-1)(k-2))$$

em que

$$b_{ij} = \sqrt{\frac{p_i(1 - p_i)}{n_i} + \frac{p_j(1 - p_j)}{n_j}}$$