

INFORME TÉCNICO SOBRE EL ASISTENTE DE MINITAB

Este documento forma parte de un conjunto de informes técnicos que explican la investigación llevada a cabo por los especialistas en estadística de Minitab para desarrollar los métodos y las verificaciones de datos que se utilizan en el Asistente de Minitab Statistical Software.

Prueba del % de defectuosos para 2 muestras

Revisión general

Una prueba para 2 proporciones se utiliza para determinar si dos proporciones difieren significativamente. En el análisis de la calidad, la prueba suele usarse cuando un producto o servicio se caracteriza por ser defectuoso o no defectuoso, para determinar si el porcentaje de elementos defectuosos difiere significativamente para muestras obtenidas de dos procesos independientes.

El Asistente de Minitab incluye una Prueba de % de defectuosos de 2 muestras. Los datos recolectados para la prueba son el número de elementos defectuosos en cada una de dos muestras independientes, que se supone que es el valor observado de una variable binomial aleatoria. El Asistente utiliza métodos exactos para calcular los resultados de las pruebas de hipótesis; por lo tanto, la tasa real de error Tipo I debe estar cerca del nivel de significancia (alfa) especificado para la prueba y no es necesaria una investigación adicional. Sin embargo, el Asistente utiliza un método de aproximación a la normal para calcular el intervalo de confianza (IC) para la diferencia en el % de defectuosos y una función de potencia teórica de la prueba de aproximación a la normal para realizar su análisis de potencia y tamaño de la muestra. Como se trata de métodos de aproximación, debemos evaluar su exactitud.

En este trabajo, investigamos las condiciones bajo las cuales los intervalos de confianza aproximados son exactos. También investigamos el método utilizado para evaluar la potencia y el tamaño de la muestra para la Prueba de % de defectuosos para 2 muestras, comparando la potencia teórica del método aproximado con la potencia real de la prueba exacta. Por último, examinamos las siguientes verificaciones de los datos que se realizan y se muestran automáticamente en la Tarjeta de informe del Asistente y explicamos de qué manera afectan los resultados del análisis:

- Validez del IC
- Tamaño de la muestra

La prueba del % de defectuosos para 2 muestras también depende de otros supuestos. Para obtener más detalles, consulte el Apéndice A.

Métodos de la prueba del % de defectuosos para 2 muestras

Exactitud de los intervalos de confianza

Aunque el Asistente utiliza la prueba exacta de Fisher para evaluar si el % de defectuosos de las dos muestras difiere significativamente, el intervalo de confianza para la diferencia se basa en el método de aproximación a la normal. Según la regla general que se encuentra en la mayoría de los libros de texto de Estadística, este intervalo de confianza aproximado es exacto si el número observado de elementos defectuosos y el número observado de elementos no defectuosos en cada muestra es por lo menos 5.

Objetivo

Queríamos evaluar las condiciones bajo las cuales son exactos los intervalos de confianza basados en la aproximación a la normal. Específicamente, queríamos determinar la manera en que la regla general relacionada con el número de elementos defectuosos y no defectuosos en cada muestra afecta la exactitud de los intervalos de confianza aproximados.

Método

La fórmula utilizada para calcular el intervalo de confianza para la diferencia entre las dos proporciones y la regla general para garantizar su exactitud se describen en el Apéndice D. Además, describimos una regla modificada y menos estricta que creamos durante el curso de nuestra investigación.

Realizamos simulaciones para evaluar la exactitud del intervalo de confianza aproximado en diferentes condiciones. Para realizar las simulaciones, generamos pares aleatorios de muestras de diferentes tamaños provenientes de varias poblaciones de Bernoulli. Para cada tipo de población de Bernoulli, calculamos un intervalo de confianza aproximado para la diferencia entre las dos proporciones en cada par de las 10,000 réplicas de muestra de Bernoulli. Posteriormente, calculamos la proporción de los 10,000 intervalos que contenía la diferencia real entre las dos proporciones, lo que se menciona como la probabilidad de cobertura simulada. Si el intervalo aproximado es exacto, la probabilidad de cobertura simulada debe estar cerca de la probabilidad de cobertura objetivo de 0.95. Para evaluar la exactitud del intervalo aproximado en lo que respecta a las reglas original y modificada para el número mínimo de elementos defectuosos y no defectuosos necesario en cada muestra, también calculamos el porcentaje de los 10,000 pares de muestras para los que se cumplió cada regla. Para obtener más detalles, consulte el Apéndice D.

Resultados

Por lo general, el intervalo de confianza aproximado para la diferencia entre dos proporciones es exacto cuando las muestras son lo suficientemente grandes, es decir, cuando el número observado de elementos defectuosos y el número observado de elementos no defectuosos en cada muestra es por lo menos 5. Por lo tanto, adoptamos esta regla para nuestra verificación de la validez del IC en la Tarjeta de informe. Aunque esta regla funciona adecuadamente en términos generales, en algunos casos puede ser demasiado conservadora y podría resultar relativamente flexible cuando las dos proporciones estén cerca de 0 ó 1. Para obtener más detalles, consulte la sección Verificaciones de los datos y el Apéndice D.

Desempeño de la función de potencia teórica

El Asistente realiza la prueba de hipótesis para comparar dos proporciones de población de Bernoulli (% de defectuosos en dos muestras) usando la prueba de Fisher. Sin embargo, puesto que la función de potencia de esta prueba exacta no se deriva fácilmente, se debe aproximar la función de potencia usando la función de potencia teórica de la prueba correspondiente de aproximación a la normal.

Objetivo

Queríamos determinar si es adecuado utilizar la función de potencia teórica basada en la prueba de aproximación a la normal para evaluar los requisitos de potencia y tamaño de la muestra para la Prueba del % de defectuosos de 2 muestras disponible en el Asistente. Para ello, tuvimos que evaluar si esta función de potencia teórica refleja fielmente la potencia real de la prueba exacta de Fisher.

Método

La metodología para la prueba exacta de Fisher, incluyendo el cálculo de su valor p , se describe en detalle en el Apéndice B. La función de potencia teórica basada en la prueba de aproximación a la normal se define en el Apéndice C. Con base en estas definiciones, realizamos simulaciones para estimar los niveles de potencia real (a los que nos referimos como niveles de potencia simulada) de la prueba exacta de Fisher cuando se utiliza para analizar la diferencia en el % de defectuosos de dos muestras.

Para realizar las simulaciones, generamos pares aleatorios de muestras de diferentes tamaños provenientes de varias poblaciones de Bernoulli. Para cada categoría de población de Bernoulli, realizamos la prueba exacta de Fisher en cada par de las 10,000 réplicas de muestra. Para cada tamaño de muestra, calculamos la potencia simulada de la prueba para detectar una diferencia determinada como la fracción de los 10,000 pares de muestras para la cual la prueba fue significativa. Para efectos de comparación, también calculamos la potencia teórica correspondiente basada en la prueba de aproximación a la normal. Si la aproximación funciona adecuadamente, los niveles de potencia teórica y simulada deberían estar cerca. Para obtener más detalles, consulte el Apéndice E.

Resultados

Nuestras simulaciones revelaron que, en general, la función de potencia teórica de la prueba de aproximación a la normal y la función de potencia simulada de la prueba exacta de Fisher son casi iguales. Por lo tanto, el Asistente utiliza la función de potencia teórica de la prueba de aproximación a la normal para estimar los tamaños de las muestras que se necesitan para detectar diferencias importantes desde el punto de vista práctico cuando se realiza la prueba exacta de Fisher.

Verificaciones de los datos

Validez del IC

Puesto que la prueba del % de defectuosos para 2 muestras utiliza una prueba exacta para evaluar la diferencia en el % de defectuosos, su exactitud no es afectada considerablemente por el número de elementos defectuosos y no defectuosos en cada muestra. Sin embargo, el intervalo de confianza para la diferencia entre los % de defectuosos se basa en una aproximación a la normal. Cuando aumenta el número de elementos defectuosos y no defectuosos en cada muestra, la exactitud del intervalo de confianza aproximado también aumenta (consulte el Apéndice D).

Objetivo



Queríamos determinar si el número de elementos defectuosos y el número de elementos no defectuosos en las muestras son suficientes para garantizar que el intervalo de confianza aproximado para la diferencia en el % de defectuosos es exacto.

Método

Utilizamos la regla general que se encuentra en la mayoría de los libros de texto de Estadística. Cuando cada muestra contiene al menos 5 elementos defectuosos y 5 elementos no defectuosos, el intervalo de confianza aproximado para la prueba del % de defectuosos para 2 muestras es exacto. Para más detalles, consulte la sección Métodos de la prueba del % de defectuosos para 2 muestras, arriba.

Resultados

Como se muestra en las simulaciones resumidas en la sección Métodos de la prueba del % de defectuosos para 2 muestras, la exactitud del intervalo de confianza depende del número mínimo de elementos defectuosos y no defectuosos en cada muestra. Por lo tanto, la Tarjeta de informe del Asistente muestra los siguientes indicadores de estado para ayudar a evaluar la exactitud del intervalo de confianza para la diferencia entre dos % de defectuosos:

Estado	Condición
	Ambas muestras tienen al menos 5 elementos defectuosos y 5 no defectuosos. El intervalo de confianza para la diferencia debería ser exacto.
	El número de elementos defectuosos o el número de elementos no defectuosos en al menos una de las muestras es inferior a 5. El intervalo de confianza para la diferencia podría no ser exacto.

Tamaño de la muestra

Normalmente, una prueba de hipótesis estadística se realiza para reunir evidencia para rechazar la hipótesis nula de que "no existe diferencia". Si la muestra es demasiado pequeña, la potencia de la prueba podría no ser adecuada para detectar una diferencia que en realidad existe, lo que produce un error Tipo II. Por lo tanto, resulta crucial asegurarse de que los tamaños de las muestras sean lo suficientemente grandes como para detectar diferencias importantes desde el punto de vista práctico con alta probabilidad.

Objetivo

Si los datos no proporcionan suficiente evidencia para rechazar la hipótesis nula, conviene determinar si los tamaños de las muestras son lo suficientemente grandes como para que la prueba detecte diferencias de interés desde el punto de vista práctico con alta probabilidad. Aunque el objetivo de la planificación del tamaño de la muestra es asegurar que los tamaños de las muestras sean lo suficientemente grandes como para detectar diferencias importantes con alta probabilidad, no deben ser tan grandes como para que las diferencias insignificantes se vuelvan estadísticamente significativas con alta probabilidad.






Método

El análisis de potencia y tamaño de la muestra para la prueba del % de defectuosos para 2 muestras se basa en la función de potencia teórica de la prueba de aproximación a la normal, que proporciona una estimación adecuada de la potencia real de la prueba exacta de Fisher (consulte los resultados de la simulación que se resumen en Desempeño de la función de potencia teórica en la sección Métodos de la prueba del % de defectuosos para 2 muestras). La función de potencia teórica puede expresarse en función de la diferencia objetivo en el % de defectuosos y el % general de defectuosos en las muestras combinadas.

Resultados

Cuando los datos no proporcionan suficiente evidencia en contra de la hipótesis nula, el Asistente utiliza la función de potencia de la prueba de aproximación a la normal para calcular las diferencias prácticas que pueden detectarse con un 80% y un 90% de probabilidad para el tamaño de muestra dado. Además, si el usuario especifica una diferencia práctica de interés en particular, el Asistente utiliza la función de potencia de la prueba de aproximación a la normal para calcular los tamaños de las muestras que producen un 80% y un 90% de probabilidad de detección de la diferencia.

Como ayuda para la interpretación de los resultados, la Tarjeta de informe del Asistente para la Prueba del % de defectuosos para 2 muestras presenta los siguientes indicadores de estado cuando se verifica la potencia y el tamaño de la muestra:

Estado	Condición
	<p>La prueba detecta una diferencia entre los % de defectuosos, así que la potencia no es un problema.</p> <p>O</p> <p>La potencia no es suficiente. La prueba no detectó una diferencia entre los % de defectuosos, pero la muestra es lo suficientemente grande como para proporcionar al menos un 90% de probabilidad de detectar la diferencia dada (potencia $\geq .90$).</p>
	<p>La potencia podría ser suficiente. La prueba no detectó una diferencia entre los % de defectuosos, pero la muestra es lo suficientemente grande como para ofrecer de un 80% a un 90% de probabilidad de detectar la diferencia dada ($.80 \leq \text{potencia} < .90$). Se indica el tamaño de la muestra que se necesita para alcanzar una potencia de 90%.</p>
	<p>La potencia podría no ser suficiente. La prueba no detectó una diferencia entre los % de defectuosos y la muestra es lo suficientemente grande como para ofrecer de un 60% a un 80% de probabilidad de detectar la diferencia dada ($.60 \leq \text{potencia} < .80$). Se indican los tamaños de las muestras que se necesitan para lograr una potencia de 80% y una potencia de 90%.</p>
	<p>La potencia no es suficiente. La prueba no detectó una diferencia entre los % de defectuosos y la muestra no es lo suficientemente grande como para proporcionar al menos un 60% de probabilidad de detectar la diferencia dada (potencia $< .60$). Se indican los tamaños de las muestras que se necesitan para lograr una potencia de 80% y una potencia de 90%.</p>
	<p>La prueba no detectó una diferencia entre los % de defectuosos. Usted no especificó una diferencia práctica que detectar. Dependiendo de sus datos, el informe puede indicar las diferencias que se podrían detectar con un 80% y un 90% de probabilidad, con base en el tamaño de la muestra y el nivel de significancia.</p>

Referencias

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

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

Apéndice A: Supuestos adicionales para la prueba del % de defectuosos para 2 muestras

La prueba del % de defectuosos para 2 muestras se basa en los siguientes supuestos:

- Los datos de cada muestra consisten en n elementos distintos, con cada elemento clasificado bien sea como defectuoso o no defectuoso.
- La probabilidad de que un elemento esté defectuoso es igual para cada elemento dentro de una muestra.
- La probabilidad de que un elemento esté defectuoso no se ve afectada por el hecho de que otro elemento esté defectuoso o no.

Estos supuestos no pueden comprobarse en las verificaciones de los datos de la Tarjeta de informe del Asistente, porque para esta prueba se ingresan datos resumidos en lugar de datos sin procesar.

Apéndice B: Prueba exacta de Fisher

Supongamos que observamos dos muestras aleatorias independientes X_1, \dots, X_{n_1} y Y_1, \dots, Y_{n_2} provenientes de distribuciones de Bernoulli, tal que

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

En las siguientes secciones, describimos los procedimientos para hacer inferencias acerca de la diferencia entre las proporciones $\delta = p_1 - p_2$.

Fórmula B1: Prueba exacta de Fisher y valor p

Para una descripción completa de la prueba exacta de Fisher, consulte Arnold (1994). En este documento, solo ofrecemos una breve descripción de la prueba.

Supongamos que V es el número de éxitos en la primera muestra y $v = n_1 \hat{p}_1$ es el número observado de éxitos en la primera muestra cuando se realiza un experimento. Supongamos además que W es el número total de éxitos en las dos muestras y que $w = n_1 \hat{p}_1 + n_2 \hat{p}_2$ son los éxitos observados cuando se realiza un experimento. Tenga en cuenta que \hat{p}_1 y \hat{p}_2 son las estimaciones de los puntos de la muestra de p_1 y p_2 .

Bajo la hipótesis nula de que $\delta = p_1 - p_2 = 0$, la distribución condicional de V dado W es la distribución hipergeométrica con la función de masa de probabilidad

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

Supongamos que $F(v|w)$ es la distribución acumulativa de probabilidad (c.d.f.) de la distribución. Entonces, los valores p para las pruebas unilateral y bilateral son los siguientes:

- **Cuando se realizan pruebas con respecto a $\delta < 0$ o equivalentemente $p_1 < p_2$**
El valor p se calcula como $F(v|w)$, donde v es el valor observado de V o el número observado de éxitos en la primera muestra y w es el valor observado de W o el número observado de éxitos en ambas muestras.
- **Cuando se realizan pruebas con respecto a $\delta > 0$ o equivalentemente $p_1 > p_2$**
El valor p se calcula como $1 - F(v-1|w)$, donde v es el valor observado de V o el número observado de éxitos en la primera muestra y w es el valor observado de W o el número observado de éxitos en ambas muestras.

- Cuando se realizan prueba con respecto a $\delta \neq 0$ o equivalentemente $p_1 \neq p_2$

El valor p se calcula según el siguiente algoritmo, donde m es el modo de distribución hipergeométrica descrito anteriormente.

- Si $v < m$, entonces el valor p se calcula como $1 - F(y - 1|w) + F(v|w)$, donde v y w son como se define arriba y $y = \min\{k \geq m: f(k|w) \leq f(v|W)\}$
- Si $v = m$, entonces el valor p es 1.0
- Si $v > m$, entonces el valor p se calcula como $1 - F(v - 1|w) + F(y|w)$, donde v y w son como se define arriba y $y = \max\{k \leq m: f(k|w) \leq f(v|W)\}$

Apéndice C: Función de potencia teórica

Para comparar dos proporciones (o, más específicamente, dos % de defectuosos), utilizamos la prueba exacta de Fisher, tal como se describe en el Apéndice B. Debido a que una función de potencia teórica de esta prueba es demasiado compleja de derivar, usamos una función de potencia aproximada. Más específicamente, utilizamos la función de potencia de la prueba de aproximación a la normal para dos proporciones, ampliamente conocida, para aproximar la potencia de la prueba exacta de Fisher.

La función de potencia de la aproximación a la normal para la prueba bilateral es

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

donde $\delta = p_1 - p_2$,

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

y $p_c = (n_1 p_1 + n_2 p_2) / (n_1 + n_2)$.

Cuando se prueba $p_1 = p_2$ en función de $p_1 > p_2$, la función de potencia es

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

Cuando se prueba $p_1 = p_2$ en función de $p_1 < p_2$, la función de potencia es

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

Apéndice D: Intervalos de confianza aproximados

Fórmula D1: Cálculo de un intervalo de confianza aproximado para la diferencia entre 2 proporciones

Un intervalo de confianza asintótico de $100(1 - \alpha)\%$ basado en la aproximación a la normal es:

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

Una regla general bien conocida para evaluar la fiabilidad de este intervalo de confianza aproximado es $n_1\hat{p}_1 \geq 5$, $n_1(1 - \hat{p}_1) \geq 5$, $n_2\hat{p}_2 \geq 5$ y $n_2(1 - \hat{p}_2) \geq 5$. En otras palabras, el intervalo de confianza es exacto si el número observado de éxitos y fracasos en cada muestra es por lo menos 5.

Nota: En esta sección y en las secciones siguientes, expresamos la regla para el intervalo de confianza en su forma más general, en términos del número de éxitos y el número de fracasos en cada muestra. Un éxito es el evento de interés y un fracaso es el complemento del evento de interés. Por lo tanto, en el contexto específico de la prueba del % de defectuosos para 2 muestras, el número de "éxitos" es equivalente al número de elementos defectuosos y el número de "fracasos" es equivalente al número de elementos no defectuosos.

Fórmula D2: Reglas para los intervalos de confianza aproximados

La regla general utilizada para los intervalos de confianza basados en la aproximación a la normal señala que los intervalos de confianza son exactos si $n_1\hat{p}_1 \geq 5$, $n_1(1 - \hat{p}_1) \geq 5$, $n_2\hat{p}_2 \geq 5$ y $n_2(1 - \hat{p}_2) \geq 5$. Es decir, el nivel de confianza real del intervalo es igual o aproximadamente igual al nivel de confianza objetivo si cada muestra contiene por lo menos 5 éxitos (defectuosos) y 5 fracasos (no defectuosos).

La regla se expresa en términos de las proporciones estimadas de éxitos y fracasos en contraposición a las proporciones reales, porque en la práctica las proporciones reales no se conocen. Sin embargo, en ámbitos teóricos donde las proporciones reales se presuponen o se conocen, la regla puede expresarse directamente en términos de las proporciones reales. En estos casos, se puede evaluar directamente de qué manera el número esperado real de éxitos y de fracasos, n_1p_1 , n_2p_2 , $n_1(1 - p_1)$ y $n_2(1 - p_2)$, afecta la probabilidad de cobertura real del intervalo de confianza para la diferencia entre las proporciones.

Podemos evaluar la probabilidad de cobertura real al seleccionar un gran número de pares de muestras con tamaños de n_1 y n_2 de las dos poblaciones de Bernoulli con las probabilidades de éxito p_1 y p_2 . Luego se calcula la probabilidad de cobertura real como la frecuencia relativa de

los pares de muestras que producen intervalos de confianza que contienen la diferencia real entre las dos proporciones. Si la probabilidad de cobertura real es exacta cuando $n_1 p_1 \geq 5$, $n_2 p_2 \geq 5$, $n_1(1 - p_1) \geq 5$ y $n_2(1 - p_2) \geq 5$, entonces según la ley fuerte de los números grandes, la probabilidad de cobertura es exacta cuando $n_1 \hat{p}_1 \geq 5$, $n_1(1 - \hat{p}_1) \geq 5$, $n_2 \hat{p}_2 \geq 5$ y $n_2(1 - \hat{p}_2) \geq 5$. Por lo tanto, cuando los niveles de confianza real y objetivo están cerca, se podría esperar que una gran proporción de los pares de las muestras provenientes de las dos poblaciones de Bernoulli fuera de tal modo que $n_1 \hat{p}_1 \geq 5$, $n_1(1 - \hat{p}_1) \geq 5$, $n_2 \hat{p}_2 \geq 5$ y $n_2(1 - \hat{p}_2) \geq 5$ si esta regla es válida. En la siguiente simulación, nos referimos a esta regla como la regla 1.

Además, en el curso de esta investigación, en muchos casos, nos dimos cuenta de que si $n_1 p_1 \geq 5$ y $n_2 p_2 \geq 5$ o si $n_1(1 - p_1) \geq 5$ y $n_2(1 - p_2) \geq 5$, entonces la probabilidad de cobertura simulada del intervalo está cerca de la cobertura objetivo. Esto condujo a una regla alternativa y más flexible que señala que los intervalos de confianza aproximados son exactos si $n_1 \hat{p}_1 \geq 5$ y $n_2 \hat{p}_2 \geq 5$ o si $n_1(1 - \hat{p}_1) \geq 5$ y $n_2(1 - \hat{p}_2) \geq 5$. En la siguiente simulación, nos referimos a esta regla modificada como la regla 2.

Simulación D1: Evaluación de la exactitud de los intervalos de confianza aproximados

Realizamos simulaciones para evaluar las condiciones bajo las cuales el intervalo de confianza aproximado para la diferencia entre dos proporciones es exacto. En particular, examinamos la exactitud del intervalo en relación con las siguientes reglas generales:

Regla 1 (original) $n_1 p_1 \geq 5$, $n_2 p_2 \geq 5$, $n_1(1 - p_1) \geq 5$ y $n_2(1 - p_2) \geq 5$

Regla 2 (modificada) $n_1 \hat{p}_1 \geq 5$ y $n_2 \hat{p}_2 \geq 5$ **O** $n_1(1 - \hat{p}_1) \geq 5$ y $n_2(1 - \hat{p}_2) \geq 5$

En cada experimento, generamos 10,000 pares de muestras de pares de poblaciones de Bernoulli definidas por las siguientes proporciones:

- **Proporciones A: tanto p_1 como p_2 están cerca de 1.0 (o cerca de 0).** Para representar este par de poblaciones de Bernoulli en la simulación, usamos $p_1 = 0.8$ y $p_2 = 0.9$.
- **Proporciones B: p_1 y p_2 están cerca de 0.5.** Para representar este par de poblaciones de Bernoulli en la simulación, usamos $p_1 = 0.4$ y $p_2 = 0.55$.
- **Proporciones C: p_1 está cerca de 0.5 y p_2 está cerca de 1.0.** Para representar este par de poblaciones de Bernoulli en la simulación, usamos $p_1 = 0.4$ y $p_2 = 0.9$.

La clasificación de proporciones descrita anteriormente se basa en la aproximación de la normal a la distribución binomial de DeMoivre-Laplace de la cual se derivan los intervalos de confianza aproximados. Se sabe que esta aproximación a la normal es exacta cuando la muestra de Bernoulli es mayor que 10 y la probabilidad de éxito está cerca de 0.5. Cuando la probabilidad de éxito está cerca de 0 ó 1, generalmente se necesita una muestra de Bernoulli más grande.

Fijamos los tamaños de las muestras para ambos pares en un único valor de n , donde $n = 10, 15, 20, 30, \dots, 100$. Limitamos el estudio a diseños balanceados ($n_1 = n_2 = n$) sin ninguna

pérdida de generalidad, porque ambas reglas dependen del número observado de éxitos y fracasos, que puede ser controlado por el tamaño de las muestras y la proporción de éxitos.

Para estimar el nivel de confianza real del intervalo de confianza para la diferencia en las dos proporciones de población (al que nos referimos como el nivel de confianza simulado), calculamos la proporción de los 10,000 intervalos que contiene la diferencia real entre las dos proporciones. La probabilidad de cobertura objetivo en cada experimento fue 0.95. Además, determinamos el porcentaje de las 10,000 muestras para el que se cumplen las condiciones conforme a las dos reglas.

Nota: Para algunas muestras pequeñas, el error estándar estimado de la diferencia entre las proporciones fue 0. Consideramos que esas muestras “se degeneran” y las descartamos del experimento. Como resultado, el número de réplicas de muestra fue un poco menos de 10,000 en algunos casos.

Los resultados se muestran en las tablas de la 1 a la 11 y gráficamente en la figura 1.

Tabla 1 Probabilidades de cobertura simulada y porcentaje de muestras que satisface la regla 1 y la regla 2 para $n=10$. La probabilidad de cobertura objetivo es 0.95.

$n = 10$							
Categoría	Proporción (p)	np	$n(1 - p)$	Probabilidad de cobertura	% de muestras que satisface la regla 1	% de muestras que satisface la regla 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			

Tabla 2 Probabilidades de cobertura simulada y porcentaje de muestras que satisface la regla 1 y la regla 2 para $n=15$. La probabilidad de cobertura objetivo es 0.95.

$n = 15$							
Categoría	Proporción (p)	np	$n(1 - p)$	Probabilidad de cobertura	% de muestras que satisface la regla 1	% de muestras que satisface la regla 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	0.930	1.2	86.9
	p_2	0.90	13.50	1.50			

Tabla 3 Probabilidades de cobertura simulada y porcentaje de muestras que satisface la regla 1 y la regla 2 para $n=20$. La probabilidad de cobertura objetivo es 0.95.

$n = 20$							
Categoría	Proporción (p)	np	$n(1 - p)$	Probabilidad de cobertura	% de muestras que satisface la regla 1	% de muestras que satisface la regla 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			

Tabla 4 Probabilidades de cobertura simulada y porcentaje de muestras que satisface la regla 1 y la regla 2 para $n=30$. La probabilidad de cobertura objetivo es 0.95.

$n = 30$							
Categoría	Proporción (p)	np	$n(1 - p)$	Probabilidad de cobertura	% de muestras que satisface la regla 1	% de muestras que satisface la regla 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			

Tabla 5 Probabilidades de cobertura simulada y porcentaje de muestras que satisface la regla 1 y la regla 2 para $n=40$. La probabilidad de cobertura objetivo es 0.95.

$n = 40$							
Categoría	Proporción (p)	np	$n(1 - p)$	Probabilidad de cobertura	% de muestras que satisface la regla 1	% de muestras que satisface la regla 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			

Tabla 6 Probabilidades de cobertura simulada y porcentaje de muestras que satisface la regla 1 y la regla 2 para $n=50$. La probabilidad de cobertura objetivo es 0.95.

$n = 50$							
Categoría	Proporción (p)	np	$n(1 - p)$	Probabilidad de cobertura	% de muestras que satisface la regla 1	% de muestras que satisface la regla 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			

Tabla 7 Probabilidades de cobertura simulada y porcentaje de muestras que satisface la regla 1 y la regla 2 para $n=60$. La probabilidad de cobertura objetivo es 0.95.

$n = 60$							
Categoría	Proporción (p)	np	$n(1 - p)$	Probabilidad de cobertura	% de muestras que satisface la regla 1	% de muestras que satisface la regla 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			

Tabla 8 Probabilidades de cobertura simulada y porcentaje de muestras que satisface la regla 1 y la regla 2 para $n=70$. La probabilidad de cobertura objetivo es 0.95.

$n = 70$						
Categoría	Proporción (p)	np	n(1 - p)	Probabilidad de cobertura	% de muestras que satisface la regla 1	% de muestras que satisface la regla 2
A	p_1	0.80	56.00	14.00	0.939	71.7
	p_2	0.90	63.00	7.00		
B	p_1	0.40	28.00	42.00	0.945	100.0
	p_2	0.55	38.50	31.50		
C	p_1	0.45	31.50	38.50	0.944	71.8
	p_2	0.90	63.00	7.00		

Tabla 9 Probabilidades de cobertura simulada y porcentaje de muestras que satisface la regla 1 y la regla 2 para $n=80$. La probabilidad de cobertura objetivo es 0.95.

$n = 80$						
Categoría	Proporción (p)	np	n(1 - p)	Probabilidad de cobertura	% de muestras que satisface la regla 1	% de muestras que satisface la regla 2
A	p_1	0.80	64.00	16.00	0.947	91.3
	p_2	0.90	72.00	8.00		
B	p_1	0.40	32.00	48.00	0.947	100.0
	p_2	0.55	44.00	36.00		
C	p_1	0.45	36.00	44.00	0.948	91.3
	p_2	0.90	72.00	8.00		

Tabla 10 Probabilidades de cobertura simulada y porcentaje de muestras que satisface la regla 1 y la regla 2 para $n=90$. La probabilidad de cobertura objetivo es 0.95.

$n = 90$						
Categoría	Proporción (p)	np	$n(1 - p)$	Probabilidad de cobertura	% de muestras que satisface la regla 1	% de muestras que satisface la regla 2
A	p_1	0.80	72.00	18.00	0.947	95.18
	p_2	0.90	81.00	9.00		
B	p_1	0.40	36.00	54.00	0.951	100.0
	p_2	0.55	49.50	40.50		
C	p_1	0.45	40.50	49.50	0.945	95.2
	p_2	0.90	81.00	9.00		

Tabla 11 Probabilidades de cobertura simulada y porcentaje de muestras que satisface la regla 1 y la regla 2 para $n=100$. La probabilidad de cobertura objetivo es 0.95.

$n = 100$						
Categoría	Proporción (p)	np	$n(1 - p)$	Probabilidad de cobertura	% de muestras que satisface la regla 1	% de muestras que satisface la regla 2
A	p_1	0.80	80.00	20.00	0.952	97.7
	p_2	0.90	90.00	10.00		
B	p_1	0.40	40.00	60.00	0.945	100.0
	p_2	0.55	55.00	45.00		
C	p_1	0.45	45.00	55.00	0.948	97.7
	p_2	0.90	90.00	10.00		

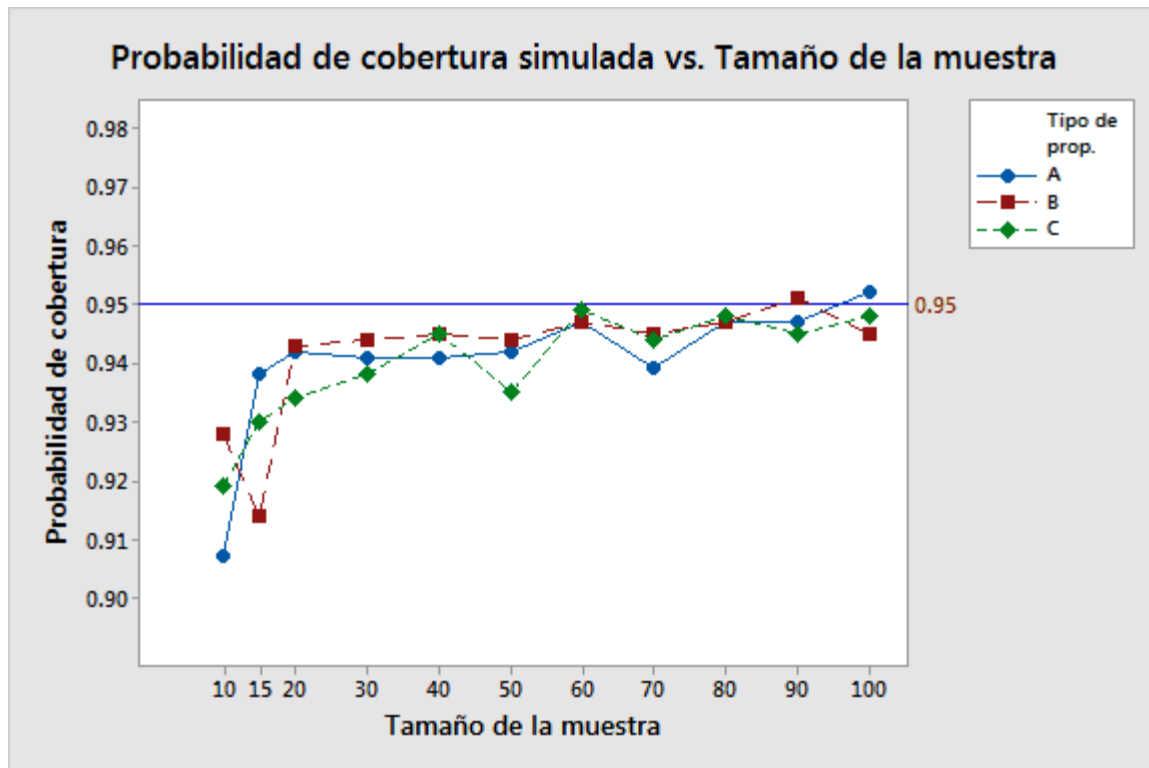


Figura 1 Probabilidades de cobertura simulada graficadas en función del tamaño de la muestra para cada categoría de las poblaciones de Bernoulli.

Los resultados de las tablas de la 1 a la 11 y la figura 1 revelan que las muestras provenientes de las poblaciones de Bernoulli en la categoría B (cuando ambas proporciones están cerca de 0.5) por lo general producen probabilidades de cobertura simulada que son más estables y están cerca de la cobertura objetivo de 0.95. En esta categoría, el número esperado de éxitos y fracasos en ambas poblaciones es mayor que en las demás categorías, incluso cuando las muestras son pequeñas.

Por otro lado, para las muestras provenientes de los pares de poblaciones de Bernoulli en la categoría A (cuando ambas proporciones están cerca de 1.0) o en la categoría C (cuando una proporción está cerca de 1.0 y la otra está cerca de 0), las probabilidades de cobertura simulada se desvían del objetivo en las muestras pequeñas, excepto cuando el número esperado de éxitos (np) o el número esperado de fracasos ($n(1-p)$) es lo suficientemente grande.

Por ejemplo, considere las muestras extraídas de las poblaciones de Bernoulli en la categoría A cuando $n = 15$. Los números esperados de éxitos son 12.0 y 13.5 y los números esperados de fracasos son 3.0 y 1.5 para cada población, respectivamente. Aunque el número esperado de fracasos es menor que 5 para ambas poblaciones, la probabilidad de cobertura simulada es de aproximadamente 0.94. Resultados como éstos nos llevaron a crear la regla 2, la cual requiere que o bien el número esperado de éxitos o bien el número esperado de fracasos sea mayor que o igual a 5 para cada muestra.

Para evaluar más a fondo qué tan eficaz es la regla 1 y la regla 2 para evaluar la aproximación para el intervalo de confianza, graficamos el porcentaje de muestras que satisfacen la regla 1 y el porcentaje de muestras que satisfacen la regla 2 en función de las probabilidades de cobertura simulada en los experimentos. Las gráficas se muestran en la figura 2.

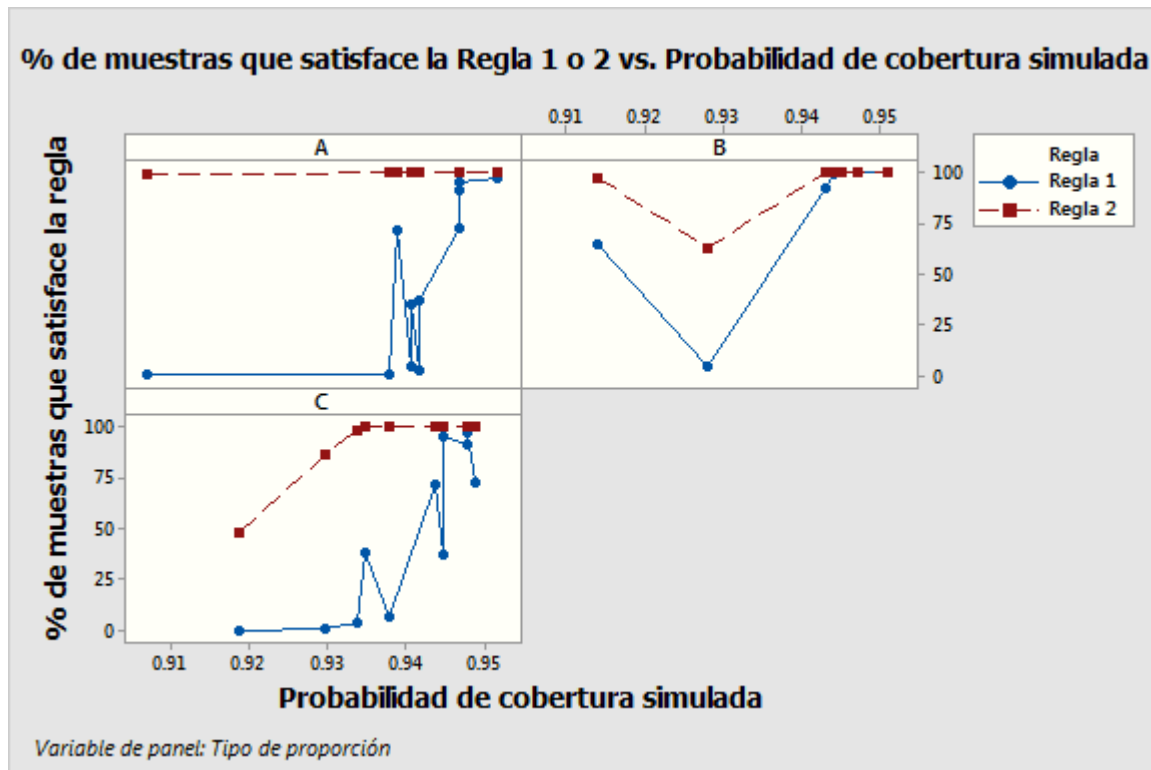


Figura 2 El porcentaje de muestras que satisfacen la regla 1 y la regla 2 graficado en función de la probabilidad de cobertura simulada, para cada categoría de las poblaciones de Bernoulli.

Las gráficas muestran que cuando las probabilidades de cobertura simulada se acercan a la cobertura objetivo de 0.95, el porcentaje de muestras que cumplen con los requisitos de cada regla por lo general se acerca a 100%. Para muestras generadas a partir de las poblaciones de Bernoulli de las categorías A y C, la regla 1 es estricta cuando las muestras son pequeñas, según lo evidenciado por el porcentaje extremadamente bajo de muestras que cumplen con la regla, aunque las probabilidades de cobertura simulada estén cerca del objetivo. Por ejemplo, cuando $n = 20$ y las muestras provienen de las poblaciones de Bernoulli de la categoría A, la probabilidad de cobertura simulada es 0.942 (véase la tabla 3). Sin embargo, la proporción de muestras que satisfacen la regla es casi 0 (0.015) (véase la figura 2). Por lo tanto, en estos casos, la regla podría ser demasiado conservadora.

La regla 2, por el contrario, es menos estricta para muestras pequeñas generadas a partir de las poblaciones de Bernoulli de la categoría A. Por ejemplo, como se muestra en la tabla 1, cuando $n = 10$ y las muestras provienen de las poblaciones de Bernoulli de la categoría A, la probabilidad de cobertura simulada es 0.907 y el 99.1% de las muestras satisface la regla.

En conclusión, la regla 1 tiende a ser excesivamente conservadora cuando las muestras son pequeñas. La Regla 2 es menos conservadora y podría ser preferible su uso cuando los tamaños de las muestras son pequeños. Sin embargo, la regla 1 es ampliamente conocida y ampliamente aceptada. Aunque la regla 2 muestra un potencial prometedor, en algunos casos puede ser demasiado liberal, como se observó anteriormente. Una posibilidad es combinar las dos reglas para aprovechar las fortalezas de cada regla; sin embargo, este enfoque requiere más investigación antes de poder ser aplicado.

Apéndice E: Comparación de la potencia real con la potencia teórica

Simulación E1: Estimación de la potencia real usando la prueba exacta de Fisher

Diseñamos una simulación para comparar los niveles de potencia real estimada (a los que nos referimos como niveles de potencia simulada) de la prueba exacta de Fisher con los niveles de potencia teórica basados en la función de potencia de la prueba de aproximación a la normal (mencionados como los niveles de potencia aproximada). En cada experimento, generamos 10,000 pares de muestras de pares de poblaciones de Bernoulli. Para cada par de muestras, las proporciones se escogieron de tal modo que la diferencia entre las proporciones fuera $p_1 - p_2 = -0.20$.

- **Proporciones A:** tanto p_1 como p_2 están cerca de 1.0 (o cerca de 0). Para representar este par de poblaciones de Bernoulli en la simulación, usamos $p_1 = 0.70$ y $p_2 = 0.90$.
- **Proporciones B:** tanto p_1 como p_2 están cerca de 0.5. Para representar este par de poblaciones de Bernoulli en la simulación, usamos $p_1 = 0.40$ y $p_2 = 0.60$.
- **Proporciones C:** p_1 está cerca de 0.5 y p_2 está cerca de 1.0. Para representar este par de poblaciones de Bernoulli en la simulación, usamos $p_1 = 0.55$ y $p_2 = 0.75$.

Fijamos los tamaños de las muestras para ambos pares en un único valor de n , donde $n = 10, 15, 20, 30, \dots, 100$. Limitamos el estudio a diseños balanceados ($n_1 = n_2 = n$), porque generalmente se parte del supuesto de que las dos muestras son del mismo tamaño. Calculamos un tamaño de muestra común necesario para detectar una diferencia importante desde el punto de vista práctico con cierta potencia.

Para estimar la potencia real para la prueba exacta de Fisher con base en los resultados de cada simulación, calculamos la fracción de los 10,000 pares de muestras para la cual la prueba bilateral fue significativa en el nivel de significancia objetivo, $\alpha = 0.05$. Posteriormente, calculamos los niveles correspondientes de potencia teórica basados en la prueba de aproximación a la normal para efectos de comparación. Los resultados se presentan en la tabla 12, a continuación.

Tabla 12 Niveles de potencia simulada de la prueba exacta de Fisher en comparación con los niveles de potencia aproximada para las tres categorías de las poblaciones de Bernoulli. El nivel de significancia objetivo es $\alpha = 0.05$.

<i>n</i>	Proporciones A		Proporciones B		Proporciones C	
	$p_1 = 0.70$ $p_2 = 0.90$		$p_1 = 0.40$ $p_2 = 0.60$		$p_1 = 0.55$ $p_2 = 0.75$	
	Potencia simulada	Potencia aproximada	Potencia simulada	Potencia aproximada	Potencia simulada	Potencia aproximada
10	0.063	0.193	0.056	0.140	0.056	0.149
15	0.151	0.271	0.097	0.190	0.101	0.204
20	0.244	0.348	0.146	0.240	0.183	0.259
30	0.370	0.490	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	0.560	0.635
70	0.814	0.849	0.610	0.661	0.649	0.703
80	0.870	0.893	0.660	0.720	0.716	0.760
90	0.907	0.925	0.716	0.770	0.772	0.808
100	0.939	0.948	0.792	0.812	0.812	0.848

Los resultados de la tabla 12 muestran que la potencia aproximada tiende a ser mayor que la potencia simulada para las tres categorías de las poblaciones de Bernoulli (A, B y C). Por ejemplo, para las proporciones de la categoría A, el tamaño de muestra real que se necesita para detectar una diferencia absoluta de -0.20 con un nivel de potencia aproximado de 0.91 es alrededor de 90. En cambio, la estimación correspondiente del tamaño de muestra basada en la función de potencia teórica aproximada es aproximadamente 85. Por lo tanto, la estimación del tamaño de muestra basada en la función de potencia aproximada suele ser ligeramente más pequeña que el tamaño de muestra real necesario para alcanzar un nivel de potencia dado.

Esta relación puede verse más claramente cuando los resultados se visualizan como curvas de potencia, como se muestra en la figura 3, abajo.

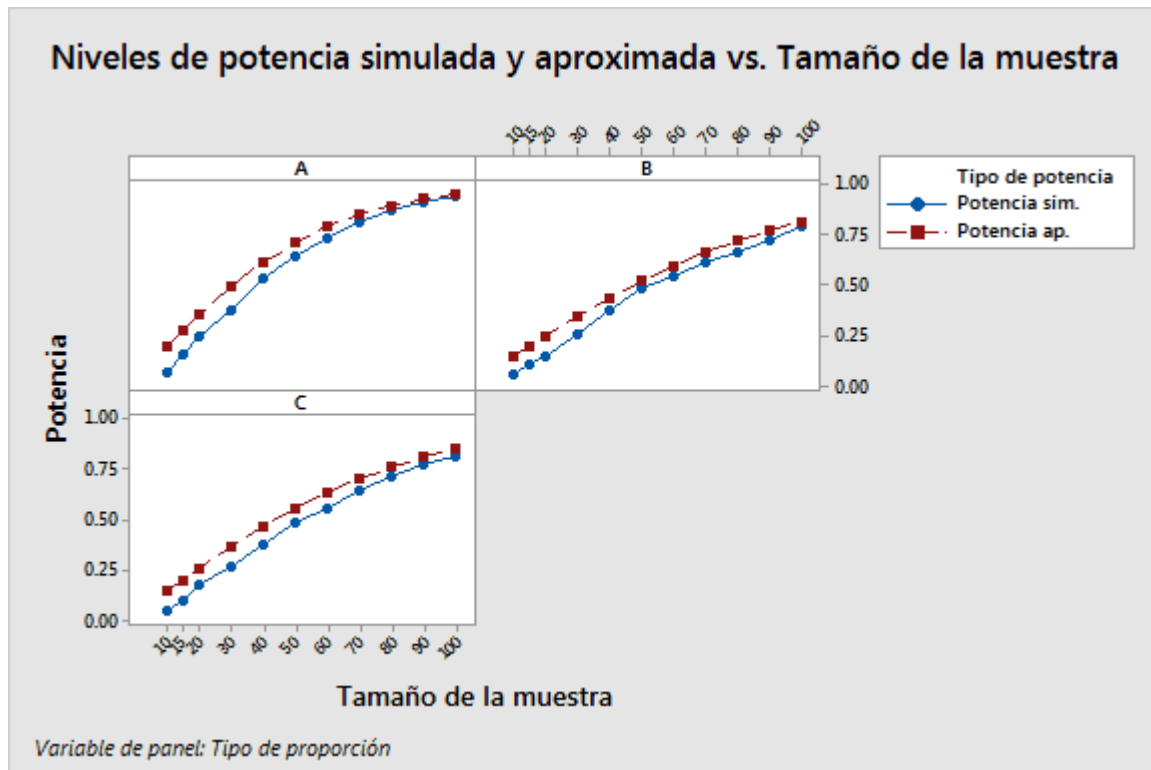


Figura 3 Gráficas de niveles de potencia simulada y aproximada de la prueba bilateral para comparar dos proporciones. Los niveles de potencia se grafican en función del tamaño de la muestra en paneles separados para cada categoría de las poblaciones de Bernoulli.

Observe que aunque las curvas de potencia simulada son más bajas que las curvas de potencia aproximada de las tres categorías de las poblaciones de Bernoulli (A, B y C), el tamaño de la diferencia entre las curvas depende de las proporciones de las poblaciones de Bernoulli de las cuales se extraen las muestras. Por ejemplo, cuando las dos proporciones están cerca de 0.5 (categoría B), los dos niveles de potencia están generalmente cerca. Sin embargo, la disparidad entre las dos curvas de potencia es más notoria en las muestras pequeñas para las proporciones asociadas con las categorías de población A y C.

Estos resultados muestran que, en general, la función de potencia teórica de la prueba de aproximación a la normal y la función de potencia simulada de la prueba exacta de Fisher son casi iguales. Por lo tanto, el Asistente utiliza la función de potencia teórica de la prueba de aproximación a la normal para estimar los tamaños de las muestras antes de realizar la prueba exacta de Fisher. Sin embargo, los tamaños de las muestras que se calculan usando la función de potencia aproximada pueden ser un poco más pequeños que los tamaños reales de las muestras que se necesitan para lograr una determinado potencia para detectar una diferencia entre las dos proporciones (% de defectuosos).

© 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.