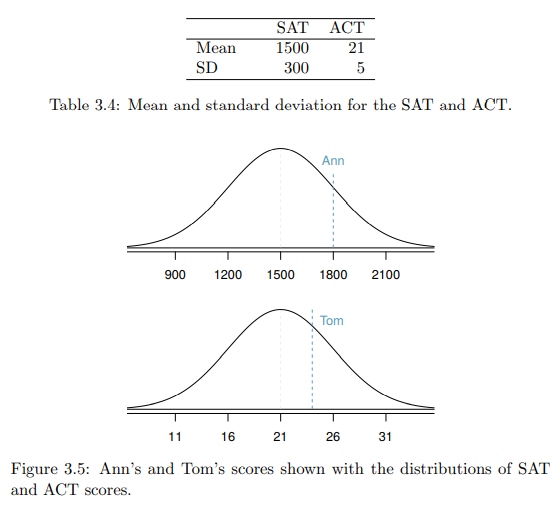
PAG 127 – 137

***3.1 Distribución normal***

***3.1.1 Modelo de distribución normal***

Muchas variables son casi normales, pero ninguna es exactamente normal. El modelo de distribución normal siempre describe una curva simétrica, unimodal y en forma de campana. Se puede ajustar usando dos parámetros: media y desviación estándar. Cambiar la media desplaza la curva de campana hacia la izquierda o hacia la derecha, mientras que el cambio de la desviación estándar estira o contrae la curva. Si una distribución normal tiene media µ y desviación estándar σ, podemos escribir la distribución como N (µ, σ). Debido a que la media y la desviación estándar describen exactamente una distribución normal, se denominan parámetros de la distribución,

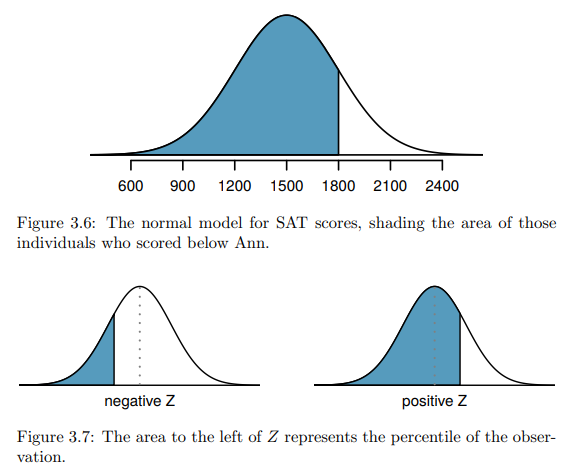


***3.1.2 Estandarización con puntajes Z***

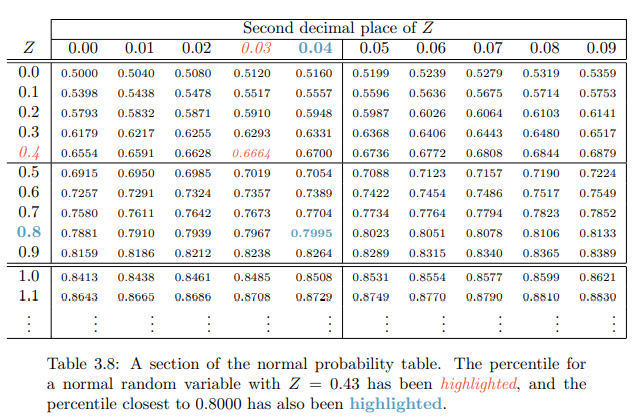
Puntuación Z de una observación se define como el número de desviaciones estándar que cae por encima o por debajo de la media. Si la observación está una desviación estándar por encima de la media, su puntuación Z es 1. Si está 1,5 desviaciones estándar por debajo de la media, entonces su puntuación Z es -1,5. Si x es una observación de una distribución N (µ, σ), definimos matemáticamente la puntuación Z como

 Z. Z-score, la observación estandarizada

Se dice que una observación x1 es más inusual que otra observación x2 si el valor absoluto de su puntuación Z es mayor que el valor absoluto de la puntuación Z de la otra observación: | Z1 | > | Z2 |. Esta técnica es especialmente útil cuando una distribución es simétrica. El área total debajo de la curva normal es siempre igual a 1.

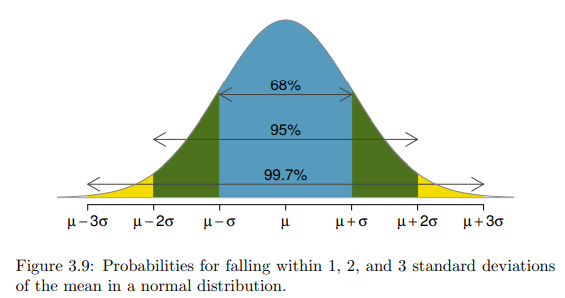


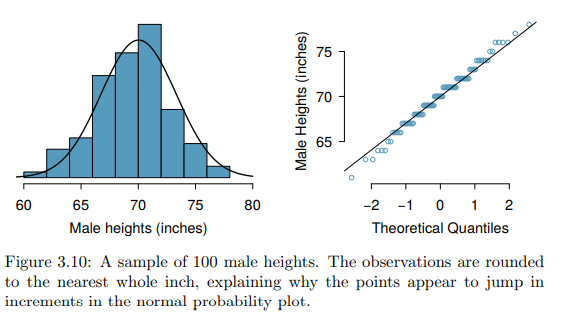
Podemos usar el modelo normal para encontrar percentiles. Se puede utilizar una tabla de probabilidad normal, que enumera las puntuaciones Z y los percentiles correspondientes, para identificar un percentil basado en la puntuación Z (y viceversa). También se puede utilizar software estadístico.



Para obtener el área hacia la derecha, cuando piden que se obtenga al menos un dato X, entonces es 1 – (percentil de Z). Es bueno dibujar la curva y sombrear el área para una estimación. También se dan los casos donde el área está dentro de un intervalo.

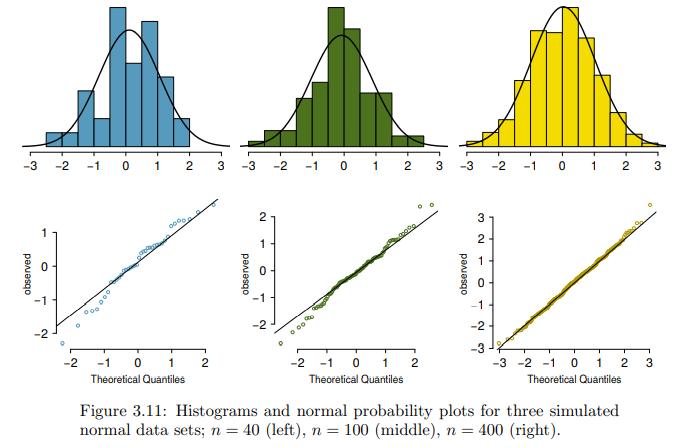
***3.1.5 Regla 68-95-99.7***





***3.2 Evaluación de la aproximación normal***

Hay dos métodos visuales para verificar el supuesto de normalidad. El primero es un histograma simple con la curva normal que mejor se ajusta superpuesta en el gráfico, como se muestra en el panel izquierdo de la Figura 3.10. La media muestral y la desviación estándar s se utilizan como parámetros de la curva normal que mejor se ajusta. Cuanto más se acerque esta curva al histograma, más razonable será la suposición del modelo normal. Otro método más común es examinar una gráfica de probabilidad normal, que se muestra en el panel derecho de la figura 3.10. Cuanto más cerca estén los puntos de una línea recta perfecta, más seguros podemos estar de que los datos siguen el modelo normal.



***3.3 Distribución geométrica (tema especial)***

***3.3.1 Distribución de Bernoulli***

Denotamos la probabilidad de éxito con p. La probabilidad de falla a veces se denota con q = 1 - p. Por lo tanto, se registra el éxito o el fracaso de cada persona en el estudio. Cuando un ensayo individual solo tiene dos resultados posibles, se denomina variable aleatoria de Bernoulli. También podemos denotar un éxito por 1 y un fracaso por 0. Además de ser conveniente para ingresar datos, también es matemáticamente útil. Supongamos que observamos diez ensayos:



Entonces, la proporción muestral, , es la media muestral de estas observaciones:



***Variable aleatoria de Bernoulli, matemática***

Si X es una variable aleatoria que toma el valor 1 con probabilidad de éxito py 0 con probabilidad 1 - p, entonces X es una variable aleatoria de Bernoulli con media y desviación estándar



***3.3.2 Distribución geométrica***

Si la probabilidad de éxito en una prueba es py la probabilidad de falla es 1 - p, entonces la probabilidad de encontrar el primer éxito en la enésima prueba viene dada por



La media (es decir, el valor esperado), la varianza y la desviación estándar de este tiempo de espera están dadas por



***3.4 Distribución binomial (tema especial)***

***3.4.1 La distribución binomial***

La distribución binomial describe la probabilidad de tener exactamente k éxitos en n ensayos de Bernoulli independientes con probabilidad de éxito p.

[# de escenarios] × P (escenario único)

Aplicamos la regla de multiplicación para eventos independientes:



Esta es nuestra fórmula general para P (escenario único). En segundo lugar, presentamos una fórmula general para el número de formas de elegir k éxitos en n ensayos, es decir, organizar k éxitos y n - k fracasos:



Suponga que la probabilidad de que una sola prueba sea exitosa es p. Entonces, la probabilidad de observar exactamente k éxitos en n ensayos independientes viene dada por



Además, la media, la varianza y la desviación estándar del número de éxitos observados son



***SUGERENCIA: ¿Es binomial? Cuatro condiciones para comprobar.***

(1) Los ensayos son independientes.

(2) El número de ensayos, n, es fijo.

(3) Cada resultado del ensayo se puede clasificar como éxito o fracaso.

(4) La probabilidad de éxito, p, es la misma para cada prueba.

***3.4.2 Aproximación normal a la distribución binomial***

La distribución binomial con probabilidad de éxito p es casi normal cuando el tamaño de la muestra n es lo suficientemente grande como para que np y n (1 - p) sean al menos 10. La distribución normal aproximada tiene parámetros correspondientes a la desviación estándar y media de la distribución binomial:



La aproximación normal se puede usar al calcular el rango de muchos éxitos posibles. Por ejemplo, podemos aplicar la distribución normal a la configuración del ejemplo 3.50.

***3.4.3 La aproximación normal se rompe en pequeños intervalos***

Se presentan diferencias en los resultados de una binomial y una normal cuando están en pequeños intervalos.

***3.5 Distribuciones más discretas (tema especial)***

***3.5.1 Distribución binomial negativa***

La distribución geométrica describe la probabilidad de observar el primer éxito en la enésima prueba. La distribución binomial negativa es más general: describe la probabilidad de observar el k-ésimo éxito en el n-ésimo ensayo.

***SUGERENCIA: ¿Es binomial negativo? Cuatro condiciones para comprobar.***

(1) Los ensayos son independientes.

(2) Cada resultado del ensayo puede clasificarse como éxito o fracaso.

(3) La probabilidad de éxito (p) es la misma para cada prueba.

(4) La última prueba debe ser un éxito.

***Distribución binomial negativa***

La distribución binomial negativa describe la probabilidad de observar el k-ésimo éxito en el n-ésimo ensayo:





Donde p es la probabilidad de que un ensayo individual sea un éxito. Se supone que todos los ensayos son independientes. En el caso binomial, normalmente tenemos un número fijo de ensayos y en su lugar consideramos el número de éxitos. En el caso del binomio negativo, examinamos cuántas pruebas se necesitan para observar un número fijo de éxitos y requerimos que la última observación sea un éxito.

***3.5.2 Distribución de Poisson***

La distribución de Poisson suele ser útil para estimar el número de eventos en una gran población durante una unidad de tiempo.Suponga que estamos atentos a los eventos y el número de eventos observados sigue una distribución de Poisson con tasa λ. Luego:



Donde k puede tomar un valor 0, 1, 2, etc., y k! representa k-factorial, como se describe en la página 146. La letra e ≈ 2.718 es la base del logaritmo natural. La desviación media y estándar de esta distribución son λ y √ λ, respectivamente. Una variable aleatoria puede seguir una distribución de Poisson si buscamos el número de eventos, la población que genera tales eventos es grande y los eventos ocurren independientemente unos de otros.

PAG 168 – 180

***Fundamentos de la inferencia***

La inferencia estadística se ocupa principalmente de comprender la calidad de las estimaciones de los parámetros. Por ejemplo, una pregunta inferencial clásica es: "¿Qué tan seguros estamos de que la media estimada, , está cerca de la media real de la población, µ?" Si bien las ecuaciones y los detalles cambian según la configuración, las bases para la inferencia son las mismas en todas las estadísticas.

**4.1 Variabilidad en estimaciones**

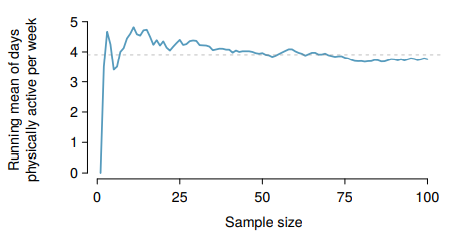
Si bien nos centramos en la media en este capítulo, las cuestiones relativas a la variación suelen ser igualmente importantes en la práctica.

***4.1.1 Estimaciones puntuales***

La media muestral se denomina estimación puntual de la media de la población: si solo podemos elegir un valor para estimar la media de la población, esta es nuestra mejor estimación. ¿Qué hay de generar estimaciones puntuales de otros parámetros de población, como la mediana de la población o la desviación estándar de la población?.

***4.1.2 Las estimaciones puntuales no son exactas***

Las estimaciones no suelen coincidir exactamente con la verdad, pero mejoran a medida que se dispone de más datos. Podemos ver esto trazando una media móvil de yrbss samp. Una media móvil es una secuencia de medias, donde cada media utiliza una observación más en su cálculo que la media directamente antes de ella en la secuencia. Por ejemplo, la segunda media de la secuencia es el promedio de las dos primeras observaciones y la tercera de la secuencia es el promedio de las tres primeras. La media móvil de la variable activa en la muestra de yrbss se muestra en la Figura 4.6 y se acerca al promedio de la población real, 3,90 días, a medida que se dispone de más datos.



*Figure 4.6: La media calculada después de agregar a cada individuo a la muestra. La media tiende a acercarse al promedio real de la población a medida que se dispone de más datos.*

Las estimaciones de los puntos muestrales solo se aproximan al parámetro de población y varían de una muestra a otra. Si tomáramos otra muestra aleatoria simple de los estudiantes de YRBSS, encontraríamos que la media muestral para el número de días activos sería un poco diferente. Será útil cuantificar qué tan variable es una estimación de una muestra a otra. Si esta variabilidad es pequeña (es decir, la media de la muestra no cambia mucho de una muestra a otra), esa estimación probablemente sea muy precisa. Si varía mucho de una muestra a otra, entonces no deberíamos esperar que nuestra estimación sea muy buena.

***4.1.3 Error estándar de la media***

*Distribución muestral*

La distribución muestral representa la distribución de las estimaciones puntuales basadas en muestras de un tamaño fijo de una determinada población. Es útil pensar en una estimación puntual particular extraída de dicha distribución. Comprender el concepto de distribución muestral es fundamental para comprender la inferencia estadística. Las medias de la muestra deberían tender a "caer alrededor" de la media de la población. Podemos ver que la media muestral tiene cierta variabilidad alrededor de la media poblacional, que se puede cuantificar usando la desviación estándar de esta distribución de medias muestrales. La desviación estándar de la media muestral nos dice qué tan lejos está la estimación típica de la media poblacional real. También describe el error típico de la estimación puntual y, por esta razón, solemos llamar a esta desviación estándar el error estándar (SE) de la estimación.

*Error estándar de una estimación*

La desviación estándar asociada con una estimación se llama error estándar. Describe el error o la incertidumbre típicos asociados con la estimación.

Dadas n observaciones independientes de una población con desviación estándar σ, el error estándar de la media muestral es igual a:



Un método confiable para garantizar que las observaciones de la muestra sean independientes es realizar una muestra aleatoria simple que consta de menos del 10% de la población. Hay un problema sutil en la ecuación: la desviación estándar de la población generalmente se desconoce. Es posible que ya haya adivinado cómo resolver este problema: podemos usar la estimación puntual de la desviación estándar de la muestra. Esta estimación tiende a ser suficientemente buena cuando el tamaño de la muestra es de al menos 30 y la distribución de la población no está muy sesgada.

**4.1.4 Propiedades básicas de las estimaciones puntuales**

Logramos tres objetivos en esta sección. Primero, determinamos que las estimaciones puntuales de una muestra pueden usarse para estimar los parámetros de la población. También determinamos que estas estimaciones puntuales no son exactas: varían de una muestra a otra. Por último, cuantificamos la incertidumbre de la media muestral utilizando lo que llamamos el error estándar, representado matemáticamente en la ecuación (4.4).

***4.2 Intervalos de confianza***

Una estimación puntual proporciona un único valor plausible para un parámetro. Sin embargo, una estimación puntual rara vez es perfecta; normalmente hay algún error en la estimación. En lugar de proporcionar solo una estimación puntual de un parámetro, el siguiente paso lógico sería proporcionar un rango plausible de valores para el parámetro.

**4.2.1 Captura del parámetro de población**

Un rango plausible de valores para el parámetro de población se denomina intervalo de confianza. Mientras más grande el rango, más probable es encontrar la respuesta.

**4.2.2 Un intervalo de confianza aproximado del 95%**

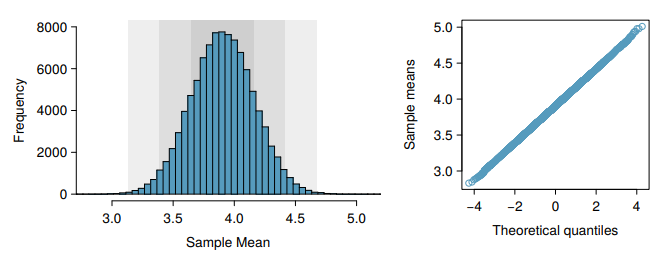
Nuestra estimación puntual es el valor más plausible del parámetro, por lo que tiene sentido construir el intervalo de confianza alrededor de la estimación puntual. El error estándar, que es una medida de la incertidumbre asociada con la estimación puntual, proporciona una guía sobre qué tan grande debería ser el intervalo de confianza. El error estándar representa la desviación estándar asociada con la estimación, y aproximadamente el 95% de las veces la estimación estará dentro de 2 errores estándar del parámetro. Si el intervalo se extiende a 2 errores estándar de la estimación puntual, podemos tener aproximadamente un 95% de confianza de que hemos capturado el parámetro verdadero:



La regla en la que aproximadamente el 95% de las observaciones se encuentran dentro de 2 desviaciones estándar de la media es solo aproximadamente cierta. Sin embargo, se mantiene muy bien para la distribución normal.

**4.2.3 La distribución muestral para la media**

En la Sección 4.1.3, presentamos una distribución muestral para , el promedio de días físicamente activos por semana para muestras de tamaño 100. Examinamos esta distribución anteriormente en la Figura 4.7. Ahora tomaremos 100.000 muestras, calcularemos la media de cada una y las trazaremos en un histograma para obtener una descripción especialmente precisa de la distribución muestral. Este histograma se muestra en el panel izquierdo de la Figura 4.9.



*Figure 4.9: El panel de la izquierda muestra un histograma de las medias muestrales para 100.000 muestras aleatorias diferentes. El panel derecho muestra una gráfica de probabilidad normal de esas medias muestrales.*

*Teorema del límite central, descripción informal*

Si una muestra consta de al menos 30 observaciones independientes y los datos no están muy sesgados, entonces la distribución de la media muestral está bien aproximada por un modelo normal.

**4.2.4 Cambiar el nivel de confianza**

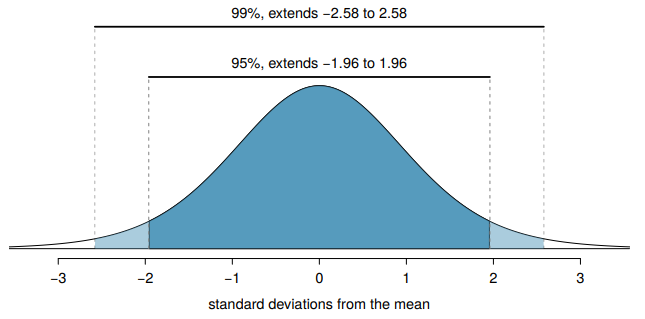
El nivel de confianza puede cambiar dependiendo de la certeza con que se quiera verificar los datos. A continuación se muestra un intervalo de confianza general del 95% para una estimación puntual que proviene de una distribución casi normal:



Hay tres componentes en este intervalo: la estimación puntual, "1,96", y el error estándar. La elección de 1,96 × SE se basó en capturar el 95% de los datos, ya que la estimación está dentro de las desviaciones estándar de 1,96 del parámetro aproximadamente el 95% del tiempo. La elección de 1,96 corresponde a un nivel de confianza del 95%.

Para crear un intervalo de confianza del 99%, cambie 1,96 en la fórmula del intervalo de confianza del 95% para que sea 2,58. La práctica guiada 4.14 destaca que el 99% de las veces una variable aleatoria normal estará dentro de 2.58 desviaciones estándar de la media. Este enfoque, que utiliza las puntuaciones Z en el modelo normal para calcular los niveles de confianza, es apropiado cuando está asociado con una distribución normal con media µ y desviación estándar . Por tanto, la fórmula para un intervalo de confianza del 99% es





*Figure 4.10: El área entre –z\* y z\* aumenta a medida que | z\* | se vuelve más grande. Si el nivel de confianza es 99%, elegimos z\* de modo que el 99% de la curva normal esté entre –z\* y z\*, que corresponde al 0,5% en la cola inferior y al 0,5% en la cola superior: z\* = 2,58.*

*Condiciones para que sea casi normal y SE sea precisa*

Condiciones importantes para ayudar a garantizar que la distribución muestral de sea casi normal y que la estimación de SE sea suficientemente precisa:

• Las observaciones de la muestra son independientes.

• El tamaño de la muestra es grande: n ≥ 30 es una buena regla general.

• La distribución de la población no está muy sesgada. Esta condición puede ser difícil de evaluar, así que use su mejor criterio.

Además, cuanto mayor sea el tamaño de la muestra, más indulgentes podemos ser con el sesgo de la muestra.

*Cómo verificar que las observaciones de la muestra sean independientes*

Si las observaciones provienen de una muestra aleatoria simple y consisten en menos del 10% de la población, entonces son independientes. Los sujetos de un experimento se consideran independientes si se les asigna aleatoriamente a los grupos de tratamiento. Si una muestra es de un proceso aparentemente aleatorio.

*Verificar un sesgo fuerte generalmente significa verificar si hay valores atípicos obvios*

Cuando existen valores atípicos prominentes, la muestra debe contener al menos 100 observaciones y, en algunos casos, mucho más.

*Intervalo de confianza para cualquier nivel de confianza*

Si la estimación puntual sigue el modelo normal con error estándar SE, entonces un intervalo de confianza para el parámetro de población es



donde z\* corresponde al nivel de confianza seleccionado.

*Margen de error*

En un intervalo de confianza, z\* × SE se denomina margen de error.

**4.2.5 Interpretación de los intervalos de confianza**

Un ojo atento podría haber observado el lenguaje algo incómodo que se usa para describir los intervalos de confianza. Interpretación correcta: Tenemos un XX% de confianza en que el parámetro de población se encuentra entre ... Un lenguaje incorrecto podría intentar describir el intervalo de confianza como capturando el parámetro de población con una cierta probabilidad. Este es un error común: si bien podría ser útil pensar en él como una probabilidad, el nivel de confianza solo cuantifica qué tan plausible es que el parámetro esté en el intervalo. Otra consideración importante de los intervalos de confianza es que solo intentan capturar el parámetro de población. Un intervalo de confianza no dice nada sobre la confianza de capturar observaciones individuales, una proporción de las observaciones o sobre la captura de estimaciones puntuales. Los intervalos de confianza solo intentan capturar los parámetros de la población.

PAG 180 – 202

***4.3 Prueba de hipótesis***

**4.3.1 Marco de prueba de hipótesis**

*Hipótesis nulas y alternativas*

La hipótesis nula () a menudo representa una perspectiva escéptica o una afirmación para ser probada. La hipótesis alternativa () representa una afirmación alternativa que se está considerando y, a menudo, está representada por un rango de posibles valores de parámetros.

La hipótesis nula a menudo representa una posición escéptica o una perspectiva sin diferencias, donde matemática mente se representa con una igualdad. La hipótesis alternativa a menudo representa una nueva perspectiva, como la posibilidad de que haya habido un cambio.

*SUGERENCIA: Marco de prueba de hipótesis*

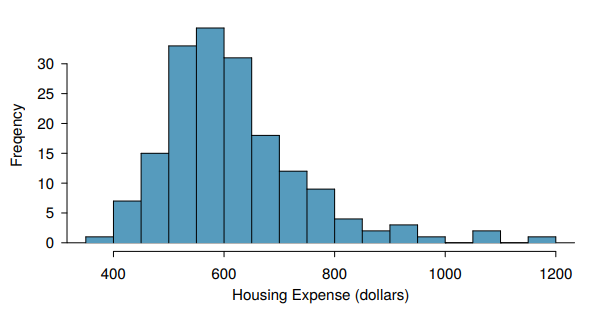
El escéptico no rechazará la hipótesis nula (), a menos que la evidencia a favor de la hipótesis alternativa () sea tan fuerte que rechace a favor de .

**4.3.2 Prueba de hipótesis utilizando intervalos de confianza**

Al calcular lo intervalos de confianza, se pueden verificar las hipótesis si los valores caen dentro o fuera de los intervalos.

*SUGERENCIA: A veces se pueden utilizar dobles negativos en las estadísticas*

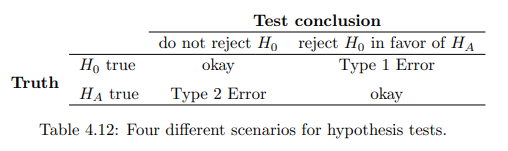
En muchas explicaciones estadísticas, utilizamos dobles negativos. Por ejemplo, podríamos decir que la hipótesis nula no es inverosímil o no rechazamos la hipótesis nula. Los dobles negativos se utilizan para comunicar que, si bien no estamos rechazando una posición, tampoco estamos diciendo que sea correcta.



*Figure 4.11: Muestra de distribución de los gastos de vivienda para estudiantes. Estos datos están muy sesgados, lo que podemos ver en la cola larga derecha con algunos valores atípicos notables*

**4.3.3 Errores de decisión**

Hay dos hipótesis en competencia: la nula y la alternativa. En una prueba de hipótesis, hacemos una declaración sobre cuál podría ser verdadera, pero podríamos elegir incorrectamente. Hay cuatro escenarios posibles, que se resumen en la Tabla 4.12.



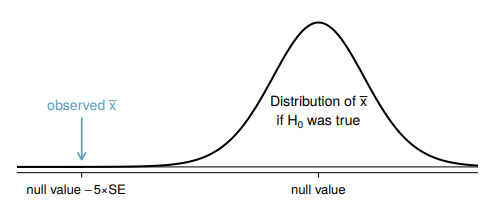
Un error de tipo 1 es rechazar la hipótesis nula cuando es realmente verdadera. Un error de tipo 2 es no rechazar la hipótesis nula cuando la alternativa es realmente cierta.

Un intervalo de confianza es, en cierto sentido, simplista en el mundo de las pruebas de hipótesis. Considere los siguientes dos escenarios:

• El valor nulo (el valor del parámetro bajo la hipótesis nula) está en el intervalo de confianza del 95% pero apenas, por lo que no rechazaríamos . Sin embargo, nos gustaría decir de alguna manera, cuantitativamente, que fue una decisión cercana.

• El valor nulo está muy fuera del intervalo, por lo que rechazamos . Sin embargo, queremos comunicar que no solo rechazamos la hipótesis nula, sino que ni siquiera se acercó. Tal caso se representa en la Figura 4.13.

En la Sección 4.3.4, presentamos una herramienta llamada **valor-p** que será útil en estos casos. El método del valor-p también se extiende a las pruebas de hipótesis en las que los intervalos de confianza no se pueden construir o aplicar fácilmente.



*Figure 4.13: Sería útil cuantificar la solidez de la evidencia contra la hipótesis nula. En este caso, la evidencia es extremadamente fuerte.*

**4.3.4 Pruebas formales usando valores-p**

El **valor-p** es una forma de cuantificar la solidez de la evidencia en contra de la hipótesis nula ya favor de la alternativa. Formalmente, el valor-p es una probabilidad condicional.

*Valor-p*

El valor-p es la probabilidad de observar datos al menos tan favorables a la hipótesis alternativa como nuestro conjunto de datos actual, si la hipótesis nula es verdadera. Normalmente utilizamos una estadística de resumen de los datos, en este capítulo la media muestral, para ayudar a calcular el valor-p y evaluar las hipótesis.

*SUGERENCIA: Pruebas unilaterales y bilaterales*

Cuando esté interesado en verificar un aumento o una disminución, pero no ambos, utilice una prueba unilateral. Cuando esté interesado en cualquier diferencia con el valor nulo (un aumento o una disminución), la prueba debe ser de dos caras.

*CONSEJO: Escriba siempre la hipótesis nula como una igualdad*

Nos resultará más útil si siempre enumeramos la hipótesis nula como una igualdad (por ejemplo, µ = 7) mientras que la alternativa siempre usa una desigualdad (por ejemplo, µ ≠ 7, µ> 7 o µ <7).

Una hipótesis nula se rechaza cuando el valor p (estadístico del valor Z calculado) es menor al nivel de significancia α = 1 – nivel de confianza.

*Valor-p como herramienta en la prueba de hipótesis*

Cuanto menor sea el valor-p, más fuertes serán los datos a favor de sobre . Un valor-p pequeño (normalmente <0,05) corresponde a evidencia suficiente para rechazar a favor de .

*SUGERENCIA: es útil hacer primero un dibujo para encontrar el valor-p*

Es útil dibujar una imagen de la distribución de como si fuera verdadera (es decir, µ es igual al valor nulo) y sombrear la región (o regiones) de las medias muestrales que son al menos tan favorables a la hipótesis alternativa. Estas regiones sombreadas representan el valor-p.

Las siguientes ideas revisan el proceso de evaluación de pruebas de hipótesis con valores-p:

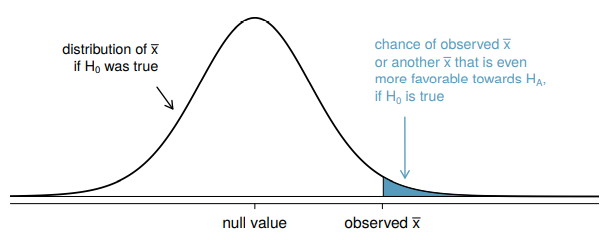
• La hipótesis nula representa la posición de un escéptico o una posición sin diferencia. Rechazamos esta posición solo si la evidencia favorece fuertemente a .

• Un valor-p pequeño significa que si la hipótesis nula es verdadera, hay una baja probabilidad de ver una estimación puntual al menos tan extrema como la que vimos. Interpretamos esto como una fuerte evidencia a favor de la alternativa.

• Rechazamos la hipótesis nula si el valor-p es menor que el nivel de significancia, α, que generalmente es 0.05. De lo contrario, no logramos rechazar .

• Siempre debemos expresar la conclusión de la prueba de hipótesis en un lenguaje sencillo para que los no estadísticos también puedan comprender los resultados.

El valor-p se construye de tal manera que podemos compararlo directamente con el nivel de significancia (α) para determinar si rechazar o no . Este método asegura que la tasa de error tipo 1 no exceda el estándar de nivel de significancia.



*Figure 4.16: Para identificar el valor-p, la distribución de la media muestral se considera como si la hipótesis nula fuera cierta. Luego, el valor-p se define y calcula como la probabilidad de la observada o una aún más favorable a bajo esta distribución.*

**4.3.5 Prueba de hipótesis de dos caras con valores-p**

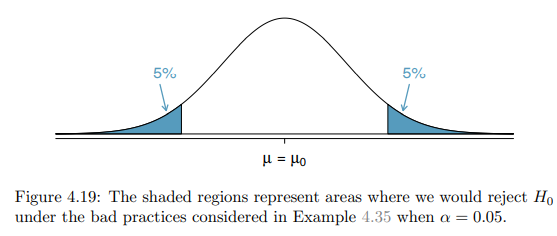
Se evalúa cuando la hipótesis alternativa es un ≠. Para rechazar la hipótesis nula, se debe cumplir que 2\*p < α.

*Precaución: las hipótesis unilaterales solo se permiten antes de ver los datos*

Después de observar los datos, es tentador convertir una prueba de dos caras en una prueba de una cara. Evite esta tentación. Las hipótesis deben establecerse antes de observar los datos. Si no es así, la prueba debe ser de dos caras.

**4.3.6 Elección de un nivel de significancia**

La elección de un nivel de significancia para una prueba es importante en muchos contextos, y el nivel tradicional es 0.05. Sin embargo, a menudo es útil ajustar el nivel de significancia según la aplicación. Podemos seleccionar un nivel que sea menor o mayor que 0.05 dependiendo de las consecuencias de las conclusiones a las que se llegue de la prueba.



Si cometer un error de tipo 1 es peligroso o especialmente costoso, deberíamos elegir un nivel de significación pequeño (por ejemplo, 0,01). En este escenario, queremos ser muy cautelosos al rechazar la hipótesis nula, por lo que exigimos pruebas muy sólidas a favor de antes de rechazar . Si un error de tipo 2 es relativamente más peligroso o mucho más costoso que un error de tipo 1, entonces deberíamos elegir un nivel de significación más alto (por ejemplo, 0,10). Aquí queremos tener cuidado de no rechazar cuando el nulo es realmente falso.

*Los niveles de importancia deben reflejar las consecuencias de los errores*

El nivel de significancia seleccionado para una prueba debe reflejar las consecuencias asociadas con los Errores Tipo 1 y Tipo 2.

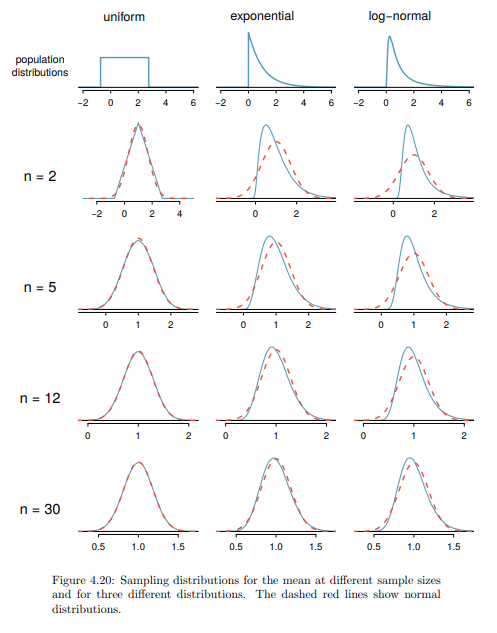
***4.4 Examen del teorema del límite central***

El modelo normal para la media muestral tiende a ser muy bueno cuando la muestra consta de al menos 30 observaciones independientes y los datos de la población no están muy sesgados. El teorema del límite central proporciona la teoría que nos permite hacer esta suposición.

*Teorema del límite central, definición informal*

La distribución de es aproximadamente normal. La aproximación puede ser pobre si el tamaño de la muestra es pequeño, pero mejora con tamaños de muestra más grandes.

El teorema del límite central establece que cuando el tamaño de la muestra es pequeño, la aproximación normal puede no ser muy buena. Sin embargo, a medida que aumenta el tamaño de la muestra, mejora la aproximación normal. Investigaremos tres casos para ver aproximadamente cuándo la aproximación es razonable. Consideramos tres conjuntos de datos: uno de una distribución uniforme, uno de una distribución exponencial y el otro de una distribución log-normal. La distribución exponencial necesitará un N más grande que la distribución uniforme para parecerse a una normal. La distribución log-normal necesitará un N más grande que la distribución exponencial para parecerse a una normal. Con N ≥ 30 para la distribución uniforme basta para que sea similar a una distribución normal.



*CONSEJO: Con n más grande, la distribución muestral de se vuelve más normal*

A medida que aumenta el tamaño de la muestra, el modelo normal para se vuelve más razonable. También podemos relajar nuestra condición de sesgo cuando el tamaño de la muestra es muy grande.

*Precaución: Examine la estructura de los datos al considerar la independencia*

Algunos conjuntos de datos se recopilan de tal manera que tienen una estructura subyacente natural entre las observaciones. Tenga especial cuidado con los supuestos de independencia con respecto a tales conjuntos de datos.

*Precaución: tenga cuidado con los valores atípicos y fuertes sesgos*

Los sesgos fuertes a menudo se identifican por la presencia de valores atípicos claros. Si un conjunto de datos tiene valores atípicos prominentes, o tales observaciones son algo comunes para el tipo de datos en estudio, entonces es útil recolectar una muestra con muchas más de 30 observaciones si el modelo normal se usará para .

***4.5 Inferencia para otros estimadores***

La media muestral no es la única estimación puntual para la que la distribución muestral es casi normal. Por ejemplo, la distribución muestral de las proporciones muestrales se parece mucho a la distribución normal cuando el tamaño de la muestra es suficientemente grande. En esta sección, presentamos varios ejemplos en los que la aproximación normal es razonable para la estimación puntual. Una estimación puntual es insesgada si la distribución muestral de la estimación se centra en el parámetro que estima. Es decir, una estimación insesgada no sobrestima ni subestima naturalmente el parámetro. Más bien, tiende a proporcionar una "buena" estimación. La media muestral es un ejemplo de una estimación puntual insesgada, al igual que cada uno de los ejemplos que presentamos en esta sección.

**4.5.1 Intervalos de confianza para estimaciones puntuales casi normales**

*Intervalo de confianza general para el caso de distribución muestral normal*

Un intervalo de confianza basado en una estimación puntual insesgada y casi normal es:

Estimación puntual ± z\*SE

Donde se selecciona z\* para que corresponda al nivel de confianza, y SE representa el error estándar. El valor z\*SE se llama margen de error.

Generalmente, el error estándar para una estimación puntual se estima a partir de los datos y se calcula mediante una fórmula. Por ejemplo, el error estándar para la media muestral es:



**4.5.2 Prueba de hipótesis para estimaciones puntuales casi normales**

*Prueba de hipótesis utilizando el modelo normal*

1. Primero escriba las hipótesis en lenguaje sencillo y luego configúrelas en notación matemática.

2. Identifique una estimación puntual adecuada del parámetro de interés.

3. Verifique las condiciones para garantizar que la estimación del error estándar sea razonable y que la estimación puntual sea casi normal e insesgada.

4. Calcule el error estándar. Haga un dibujo que represente la distribución de la estimación bajo la idea de que es verdadera. Áreas de sombra que representan el valor-p.

5. Utilizando la imagen y el modelo normal, calcule la estadística de prueba (puntuación Z) e identifique el valor-p para evaluar las hipótesis. Escribe una conclusión en un lenguaje sencillo.

La puntuación Z se llama estadística de prueba. En la mayoría de las pruebas de hipótesis, una estadística de prueba es un resumen de datos particular que es especialmente útil para calcular el valor-p y evaluar la prueba de hipótesis. En el caso de estimaciones puntuales que son casi normales, la estadística de prueba es la puntuación Z.

*Estadística de prueba*

Una estadística de prueba es una estadística de resumen que es particularmente útil para evaluar una prueba de hipótesis o identificar el valor-p. Cuando una estimación puntual es casi normal, utilizamos la puntuación Z de la estimación puntual como estadística de prueba. En capítulos posteriores encontramos situaciones en las que otras estadísticas de prueba son útiles.

**4.5.3 Estimaciones puntuales no normales**

Podemos aplicar las ideas de los intervalos de confianza y las pruebas de hipótesis a los casos en los que la estimación puntual o la estadística de prueba no son necesariamente normales. Hay muchas razones por las que puede surgir tal situación:

• El tamaño de la muestra es demasiado pequeño para que la aproximación normal sea válida;

• La estimación del error estándar puede ser deficiente; o

• La estimación puntual tiende a una distribución que no es la distribución normal.

Para cada caso en el que la aproximación normal no es válida, nuestra primera tarea siempre es comprender y caracterizar la distribución muestral de la estimación puntual o estadística de prueba. A continuación, podemos aplicar los marcos generales para los intervalos de confianza y la prueba de hipótesis a estas distribuciones alternativas.

**4.5.4 Cuándo retirarse**

Las herramientas estadísticas dependen de las condiciones. Cuando no se cumplen las condiciones, estas herramientas no son confiables y sacar conclusiones de ellas es traicionero. Las condiciones para estas herramientas suelen presentarse de dos formas.

• Las observaciones individuales deben ser independientes. Una muestra aleatoria de menos del 10% de la población asegura que las observaciones sean independientes. En los experimentos, generalmente requerimos que los sujetos sean asignados al azar en grupos. Si falla la independencia, se deben utilizar técnicas avanzadas y, en algunos casos, la inferencia puede no ser posible.

• Otras condiciones se centran en el tamaño de la muestra y el sesgo. Por ejemplo, si el tamaño de la muestra es demasiado pequeño, el sesgo es demasiado fuerte o existen valores atípicos extremos, el modelo normal para la media muestral fallará.

Siempre es necesaria la verificación de las condiciones de las herramientas estadísticas. Siempre que no se cumplan las condiciones para una técnica estadística, existen tres opciones. El primero es aprender nuevos métodos que sean apropiados para los datos. La segunda ruta es consultar a un estadístico. La tercera ruta es ignorar la falla de las condiciones. Esta última opción invalida efectivamente cualquier análisis y puede desacreditar hallazgos novedosos e interesantes.

**4.5.5 Significación estadística versus significación práctica**

Cuando el tamaño de la muestra aumenta, las estimaciones puntuales se vuelven más precisas y cualquier diferencia real en el valor medio y nulo se vuelve más fácil de detectar y reconocer. Centrarse en encontrar un resultado significativo. Tener el tamaño de mientras mínimo ideal y una estimación razonable de la desviación estándar, como también un nivel de significancia.

PAG 219 – 239

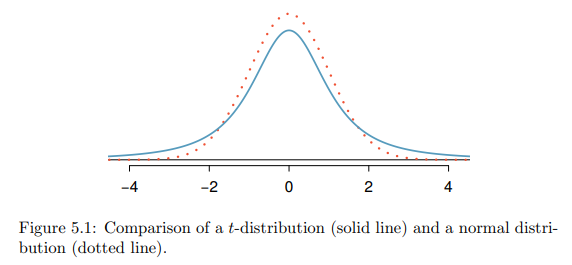
***5.1 Medias de una muestra con distribución-t***

**5.1.1 La condición de normalidad**

Un caso especial del teorema del límite central asegura que la distribución de las medias muestrales será casi normal, independientemente del tamaño de la muestra, cuando los datos provienen de una distribución casi normal.

*Teorema del límite central para datos normales*

La distribución muestral de la media es casi normal cuando las observaciones muestrales son independientes y provienen de una distribución casi normal. Esto es cierto para cualquier tamaño de muestra.



Si bien este parece un caso especial muy útil, hay un pequeño problema. Es intrínsecamente difícil verificar la normalidad en pequeños conjuntos de datos.

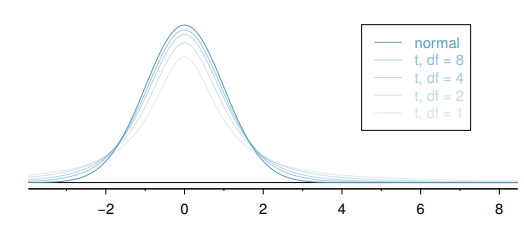
*Precaución: Comprobación del estado de normalidad*

Debemos tener precaución al verificar la condición de normalidad para muestras pequeñas. Es importante no solo examinar los datos, sino también pensar de dónde provienen. Por ejemplo, pregunte: ¿esperaría que esta distribución fuera simétrica y estoy seguro de que los valores atípicos son raros?

Puede relajar la condición de normalidad a medida que aumenta el tamaño de la muestra. Si el tamaño de la muestra es 10 o más, una ligera desviación no es problemática. Una vez que el tamaño de la muestra llega a alrededor de 30, entonces un sesgo moderado es razonable. Los datos con un fuerte sesgo o valores atípicos requieren un análisis más cauteloso.

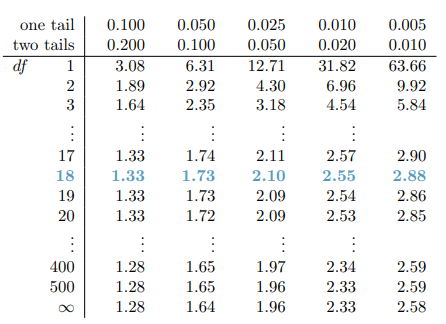
**5.1.2 Introducción de la distribución-t**

En los casos en los que usaremos una muestra pequeña para calcular el error estándar, será útil confiar en una nueva distribución para los cálculos de inferencia: la distribución-t. Una distribución-t tiene forma de campana. Sin embargo, sus colas son más gruesas que las del modelo normal. Esto significa que es más probable que las observaciones caigan más allá de dos desviaciones estándar de la media que por debajo de la distribución normal. Dan la corrección que necesitamos para resolver el problema de un error estándar mal estimado. La distribución-t, siempre centrada en cero, tiene un solo parámetro: grados de libertad. Los grados de libertad (df) describen la forma precisa de la distribución-t en forma de campana.

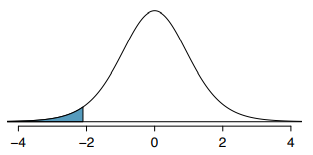


*Figure 5.2: Cuanto mayores son los grados de libertad, más se asemeja la distribución-t al modelo normal estándar.*

Cuando los grados de libertad son aproximadamente 30 o más, la distribución-t es casi indistinguible de la distribución normal. Es muy útil familiarizarse con la distribución-t, porque nos permite una mayor flexibilidad que la distribución normal al analizar datos numéricos.



*Table 5.3: Una mirada abreviada a la tabla-t. Cada fila representa una distribución-t diferente. Las columnas describen los límites para áreas específicas de la cola. La fila con df = 18 se ha resaltado.*



*Figure 5.4: La distribución-t con 18 grados de libertad. El área por debajo de -2,10 se ha sombreado.*

**5.1.3 Condiciones para usar la distribución-t para la inferencia en una media muestral**

Para continuar con la distribución-t para la inferencia sobre una sola media, primero verificamos dos condiciones.

**Independencia de observaciones**. Recopilamos una muestra aleatoria simple de menos del 10% de la población, o si los datos son de un experimento o proceso aleatorio, verificamos lo mejor que podamos que las observaciones sean independientes.

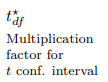
**Las observaciones provienen de una distribución casi normal**. Esta segunda condición es difícil de verificar con pequeños conjuntos de datos. A menudo (i) echamos un vistazo a una gráfica de los datos para detectar desviaciones obvias del modelo normal y (ii) consideramos si alguna experiencia previa nos alerta de que los datos pueden no ser casi normales.

*SUGERENCIA: Cuándo usar la distribución-t*

Utilice la distribución-t para la inferencia de la media muestral cuando las observaciones son independientes y casi normales. Puede relajar la condición casi normal a medida que aumenta el tamaño de la muestra. Por ejemplo, la distribución de datos puede estar moderadamente sesgada cuando el tamaño de la muestra es de al menos 30.

5.1.4 Intervalos de confianza-t de una muestra

En el modelo normal, usamos z\* y el error estándar para determinar el ancho de un intervalo de confianza. Revisamos ligeramente la fórmula del intervalo de confianza cuando usamos la distribución-t:

La media de la muestra y el error estándar estimado se calculan como antes. El valor es un punto de corte que obtenemos con base en el nivel de confianza y la distribución t con grados de libertad df. Antes de determinar este límite, primero necesitaremos los grados de libertad.

*Grados de libertad para una sola muestra*

Si la muestra tiene n observaciones y estamos examinando una sola media, entonces usamos la distribución-t con df = n - 1 grados de libertad.

*Encontrar un intervalo de confianza-t para la media*

Basado en una muestra de n observaciones independientes y casi normales, un intervalo de confianza para la media poblacional es



Donde es la media de la muestra, corresponde al nivel de confianza y grados de libertad, y SE es el error estándar estimado por la muestra.

**5.1.5 Pruebas-t de una muestra**

Con la independencia satisfecha y un ligero sesgo que no es una preocupación para una muestra tan grande, podemos proceder a realizar una prueba de hipótesis usando la distribución-t.

*Cuando usamos una distribución t, usamos una puntuación-T (igual que la puntuación Z)*

Para ayudarnos a recordar usar la distribución-t, usamos una T para representar la estadística de prueba, y a menudo lo llamamos puntaje-T. El puntaje Z y el puntaje-T se calculan exactamente de la misma manera y son conceptualmente idénticos: cada uno representa cuántos errores estándar es el valor observado del valor nulo.

***5.2 Datos emparejados***

*Datos emparejados*

Se emparejan dos conjuntos de observaciones si cada observación de un conjunto tiene una correspondencia o conexión especial con exactamente una observación del otro conjunto de datos.

Para analizar datos emparejados, a menudo es útil observar la diferencia en los resultados de cada par de observaciones.

**5.2.2 Inferencia para datos emparejados**

Para analizar un conjunto de datos emparejados, simplemente analizamos las diferencias. Podemos utilizar las mismas técnicas de distribución t que aplicamos en la última sección.

***5.3 Diferencia de dos medias***

En esta sección consideramos una diferencia en dos medias poblacionales, − , bajo la condición de que los datos no estén emparejados. Al igual que con una sola muestra, identificamos condiciones para asegurarnos de que podemos usar la distribución t con una estimación puntual de la diferencia, - .

**5.3.1 Intervalo de confianza para una diferencia de medias**

*Usando la distribución-t para una diferencia de medias*

La distribución-t se puede usar para inferencia cuando se trabaja con la diferencia estandarizada de dos medias si (1) cada muestra cumple las condiciones para usar la distribución-t y (2) las muestras son independientes.

*Distribución de una diferencia de medias muestrales*

La diferencia muestral de dos medias, ¯x1 - x¯2, se puede modelar utilizando la distribución-t y el error estándar



cuando cada media de la muestra se puede modelar por sí misma usando una distribución-t y las muestras son independientes. Para calcular los grados de libertad, utilice software estadístico o el menor de - 1 y - 1.

**5.3.2 Pruebas de hipótesis basadas en una diferencia de medias**

Al igual que en casos anteriores, se pueden ocupar las diferencias de medias para calcular las estimaciones puntuales

**5.3.4 Resumen para inferencia usando la distribución-t**

**Pruebas de hipótesis**. Al aplicar la distribución-t para una prueba de hipótesis, procedemos de la siguiente manera:

• Escriba hipótesis adecuadas.

• Verifique las condiciones para usar la distribución-t.

- Una muestra o diferencias de datos emparejados: las observaciones (o diferencias) deben ser independientes y casi normales. Para tamaños de muestra más grandes, podemos relajar el requisito casi normal, p. Ej. un ligero sesgo está bien para tamaños de muestra de 15, un sesgo moderado para tamaños de muestra de 30 y un sesgo fuerte para tamaños de muestra de 60.

- Para una diferencia de medias cuando los datos no están emparejados: cada media muestral debe satisfacer por separado las condiciones de una muestra para la distribución-t, y los datos de los grupos también deben ser independientes.

• Calcule la estimación puntual de interés, el error estándar y los grados de libertad. Para df, use n - 1 para una muestra, y para dos muestras use software estadístico o el menor de - 1 y - 1.

• Calcule la puntuación-T y el valor-p.

• Haga una conclusión basada en el valor-p y escriba una conclusión en contexto y en un lenguaje sencillo para que cualquiera pueda entender el resultado.

**Intervalos de confianza**. De manera similar, lo siguiente es cómo calculamos generalmente un intervalo de confianza usando una distribución-t:

• Verifique las condiciones para usar la distribución-t. (Véase más arriba.)

• Calcule la estimación puntual de interés, el error estándar, los grados de libertad y .

• Calcule el intervalo de confianza utilizando la fórmula general, estimación puntual ± SE.

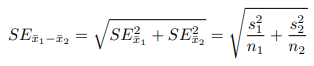
• Ponga las conclusiones en contexto y en un lenguaje sencillo para que incluso los no estadísticos puedan comprender los resultados.

**5.3.5 Examen de la fórmula del error estándar (tema especial)**

La fórmula para el error estándar de la diferencia en dos medias es similar a la fórmula para otros errores estándar. Recuerde que el error estándar de una sola media, ¯x1, se puede aproximar mediante



donde y representan la desviación estándar de la muestra y el tamaño de la muestra. El error estándar de la diferencia de dos medias muestrales se puede construir a partir de los errores estándar de las medias muestrales separadas:



Esta relación especial se deriva de la teoría de la probabilidad.

**5.3.6 Estimación de la desviación estándar agrupada (tema especial)**

La **desviación estándar combinada** de dos grupos es una forma de utilizar los datos de ambas muestras para estimar mejor la desviación estándar y el error estándar. Si y son las desviaciones estándar de los grupos 1 y 2 y hay buenas razones para creer que las desviaciones estándar de la población son iguales, entonces podemos obtener una estimación mejorada de las varianzas de grupo al combinar sus datos:



donde y son los tamaños de muestra, como antes. Para usar esta nueva estadística, sustituimos s2 agrupado en lugar de y en la fórmula de error estándar, y usamos una fórmula actualizada para los grados de libertad:



Los beneficios de combinar la desviación estándar se obtienen mediante la obtención de una mejor estimación de la desviación estándar para cada grupo y el uso de un parámetro de grados de libertad más grande para la distribución-t. Ambos cambios pueden permitir un modelo más preciso de la distribución muestral de − , si las desviaciones estándar de los dos grupos son iguales.

*Precaución: agrupe las desviaciones estándar solo después de una cuidadosa consideración*

Una desviación estándar combinada solo es apropiada cuando la investigación de antecedentes indica que las desviaciones estándar de la población son casi iguales. Cuando el tamaño de la muestra es grande y la condición puede verificarse adecuadamente con datos, los beneficios de combinar las desviaciones estándar disminuyen enormemente.

***5.4 Cálculos de poder para una diferencia de medias (tema especial)***

A menudo, en la planificación de experimentos, hay dos consideraciones en competencia:

• Queremos recopilar datos suficientes para poder detectar efectos importantes.

• La recopilación de datos puede resultar costosa y, en experimentos con personas, puede haber algún riesgo para los pacientes.

En esta sección, nos enfocamos en el contexto de un ensayo clínico, que es un experimento relacionado con la salud donde el sujeto son personas, y determinaremos un tamaño de muestra apropiado donde podemos estar 80% seguros de que detectaremos cualquier efecto prácticamente importante.

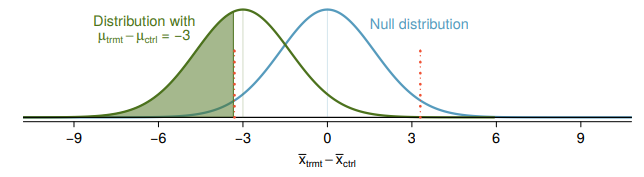
PAG 239 – 245

**5.4.1 Realización de los movimientos de una prueba**

Vamos a pasar por los movimientos de una prueba de hipótesis. Esto nos ayudará a enmarcar nuestros cálculos para determinar un tamaño de muestra apropiado para el estudio. A continuación, realizaremos algunos cálculos hipotéticos para determinar la probabilidad de que rechacemos la hipótesis nula, si la hipótesis alternativa fuera realmente cierta.

**5.4.2 Calcular la poder para una prueba de 2 muestras**

Al planificar un estudio, queremos saber la probabilidad de que detectemos un efecto que nos importa. En otras palabras, si hay un efecto real y ese efecto es lo suficientemente grande como para tener un valor práctico, ¿cuál es la probabilidad de que detectemos ese efecto? Esta probabilidad se llama poder y podemos calcularla para diferentes tamaños de muestra o para diferentes tamaños de efecto. Primero determinamos cuál es un resultado prácticamente significativo. Por ejemplo, cuando se tiene una muestra y se calcula su intervalo de confianza. Luego el valor nulo cambia, de crea otra distribución normal con el nuevo valor nulo, posicionándose en el lado correcto de la distribución original, por ejemplo:



Luego calculamos el puntaje Z con el nuevo valor nulo y con el valor estimado será el extremo del intervalo de confianza original más cercano. El resultado equivale al poder, o a la probabilidad de rechazar la hipótesis nula a favor de la alternativa.

**5.4.3 Determinación de un tamaño de muestra adecuado**

Para esto, dado un poder, se debe calcular el valor de la puntuación Z usando el valor del poder y luego sumarlo con la puntuación Z del nivel de confianza, multiplicando el resultado por SE. Este resultado equivale a la distancia y debe ser igualado al tamaño de efecto de interés. Luego, usando la fórmula de SE, se puede despejar n para obtener el tamaño mínimo ideal.

En el último ejemplo, encontramos que si tenemos un tamaño de muestra de 100 en cada grupo, solo podemos detectar un tamaño de efecto de 3 mmHg con una probabilidad de aproximadamente 0,42. Supongamos que los investigadores avanzaron y solo usaron 100 pacientes por grupo, y los datos no respaldaron la hipótesis alternativa, es decir, los investigadores no rechazaron . Esta es una situación muy mala por algunas razones:

• En el fondo de la mente de los investigadores, todos se estarían preguntando, tal vez haya una diferencia real y significativa, pero no pudimos detectarla con una muestra tan pequeña.

• La empresa probablemente invirtió cientos de millones de dólares en el desarrollo del nuevo fármaco, por lo que ahora tienen una gran incertidumbre sobre su poder, ya que el experimento no tuvo una gran oportunidad de detectar efectos que aún podrían ser importantes.

• Los pacientes fueron sometidos a la droga, y ni siquiera podemos decir con mucha certeza que la droga no ayuda (o daña) a los pacientes.

• Es posible que sea necesario realizar otro ensayo clínico para obtener una respuesta más concluyente sobre si el medicamento tiene algún valor práctico, y realizar un segundo ensayo clínico puede llevar años y muchos millones de dólares.

Queremos evitar esta situación, por lo que debemos determinar un tamaño de muestra adecuado para asegurarnos de que podemos estar bastante seguros de que detectaremos cualquier efecto que sea prácticamente importante. Como se mencionó anteriormente, se consideró que un cambio de 3 mmHg era la diferencia mínima que era prácticamente importante. Como primer paso, podríamos calcular el poder para varios tamaños de muestra diferentes. Por ejemplo, probemos con 500 pacientes por grupo.

Los cálculos de energía para experimentos costosos o arriesgados son críticos. Sin embargo, ¿qué pasa con los experimentos que son económicos y donde las consideraciones éticas son mínimas? Por ejemplo, si estamos haciendo la prueba final en una nueva función en un sitio web popular, ¿cómo cambiarían nuestras consideraciones sobre el tamaño de la muestra? Como antes, queremos asegurarnos de que la muestra sea lo suficientemente grande. Sin embargo, si la función se ha sometido a algunas pruebas y se sabe que funciona bien (es decir, no frustra a muchos usuarios del sitio), es posible que realicemos un experimento mucho más grande de lo necesario para detectar los efectos mínimos de interés. La razón es que puede haber beneficios adicionales al tener una estimación aún más precisa del efecto de la nueva función. Incluso podemos realizar un gran experimento como parte del lanzamiento de la nueva función.

PAG 274 – 286

***CAPÍTULO 6: INFERENCIA PARA DATOS CATEGÓRICOS***

El capítulo 6 introduce la inferencia en el contexto de datos categóricos. Los métodos que aprendimos en capítulos anteriores seguirán siendo útiles en estos entornos. Por ejemplo, las proporciones de la muestra están bien caracterizadas por una distribución casi normal cuando se satisfacen ciertas condiciones, lo que hace posible emplear el intervalo de confianza habitual y las herramientas de prueba de hipótesis. En otros casos, como aquellos con tablas de contingencia o cuando no se cumplen las condiciones del tamaño de la muestra, usaremos una distribución diferente, aunque las ideas centrales siguen siendo las mismas.

***6.1 Inferencia para una sola proporción***.

**6.1.1 Identificar cuándo la proporción de la muestra es casi normal**

Una proporción muestral puede describirse como media muestral. Si representamos cada "éxito" como un 1 y cada "fracaso" como un 0, entonces la proporción de la muestra es la media de estos resultados numéricos:



La distribución de es casi normal cuando la distribución de 0 y 1 no está demasiado sesgada para el tamaño de la muestra. La pauta más común para el tamaño de la muestra y el sesgo cuando se trabaja con proporciones es garantizar que esperamos observar un número mínimo de éxitos (1) y fracasos (0), generalmente al menos 10 de cada uno. Las etiquetas de éxito y fracaso no tienen por qué significar algo positivo o negativo. Estos términos son solo palabras convenientes que se usan con frecuencia cuando se habla de proporciones.

*Condiciones para que la distribución muestral de sea casi normal*

La distribución muestral para , tomada de una muestra de tamaño n de una población con una proporción verdadera p, es casi normal cuando las observaciones de la muestra son independientes y esperábamos ver al menos 10 éxitos y 10 fracasos en nuestra muestra , es decir, np ≥ 10 yn (1 - p) ≥ 10. Esto se denomina condición de éxito-fracaso. Si se cumplen estas condiciones, entonces la distribución muestral de es casi normal con media py error estándar:



= proporción de la muestra, = proporción de la población.

*SUGERENCIA: recordatorio sobre cómo comprobar la independencia de las observaciones*

Si los datos provienen de una muestra aleatoria simple y consisten en menos del 10% de la población, entonces el supuesto de independencia es razonable. Alternativamente, si los datos provienen de un proceso aleatorio, debemos evaluar la condición de independencia con más cuidado.

**6.1.2 Intervalos de confianza para una proporción**

El cálculo de los intervalos de confianza es el mismo que para las distribuciones Z, solo que esta vez se usa el SE para las proporciones.

*Construir un intervalo de confianza para una proporción*

• Verifique que las observaciones sean independientes y también verifique la condición de éxito-fracaso usando y n.

• Si se cumplen las condiciones, la distribución muestral de puede ser bien aproximada por el modelo normal.

• Construya el error estándar usando en lugar de p y aplique la fórmula del intervalo de confianza general.

**6.1.3 Prueba de hipótesis para una proporción**

Para aplicar el marco de distribución normal en el contexto de una prueba de hipótesis para una proporción, se deben satisfacer las condiciones de independencia y éxito-fracaso. En una prueba de hipótesis, la condición de éxito-fracaso se verifica usando la proporción nula: verificamos que y n (1 - ) son al menos 10, donde es el valor nulo.

Se calcula la puntuación Z donde el valor estimado y el valor nulo son proporciones, donde si el valor nulo no se define, entonces por defecto es 0.5.

*Prueba de hipótesis para una proporción*

Establezca hipótesis y verifique las condiciones utilizando el valor nulo, , para asegurarse de que sea casi normal en . Si las condiciones se cumplen, construya el error estándar, nuevamente usando , y muestre el valor-p en un dibujo. Por último, calcule el valor-p y evalúe las hipótesis.

**6.1.4 Elección de un tamaño de muestra al estimar una proporción**

El margen de error para una proporción de muestra es:



Esta fórmula se debe comparar, por igualdad o inecuaciones, con un margo de error estimado. Si no hay estimación p, se puede usar por defecto 0.5, por tanto la única incógnita es *n*.

***6.2 Diferencia de dos proporciones***

Nos gustaría sacar conclusiones sobre la diferencia en dos proporciones de población: - . En nuestras investigaciones, primero identificamos una estimación puntual razonable de - basada en la muestra. Puede que ya hayas adivinado su forma: - . A continuación, verificamos que la estimación puntual sigue el modelo normal al verificar ciertas condiciones. Finalmente, calculamos el error estándar de la estimación y aplicamos nuestro marco inferencial.

**6.2.1 Distribución muestral de la diferencia de dos proporciones**

Debemos comprobar dos condiciones antes de aplicar el modelo normal a - . Primero, la distribución muestral para cada proporción de muestra debe ser casi normal y, segundo, las muestras deben ser independientes. En estas dos condiciones, la distribución muestral de - puede aproximarse bien utilizando el modelo normal.

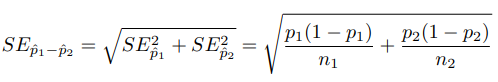
*Condiciones para que la distribución muestral de - sea normal*

La diferencia - tiende a seguir un modelo normal cuando:

•Cada proporción sigue por separado un modelo normal, y

•Las dos muestras son independientes entre sí.

El error estándar de la diferencia en las proporciones muestrales es:



Donde y representan las proporciones de la población, y y representan los tamaños de muestra.

Para la diferencia en dos medias, la fórmula de error estándar tomó la siguiente forma:



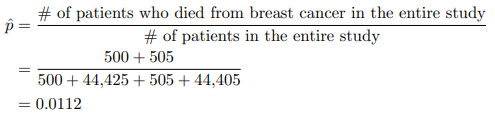
El error estándar para la diferencia en dos proporciones toma una forma similar. Las razones detrás de esta similitud tienen su origen en la teoría de la probabilidad de la Sección 2.4, que se describe para este contexto en la Práctica guiada 5.28 en la página 238.

**6.2.2 Intervalos de confianza para y**

En el establecimiento de intervalos de confianza para una diferencia de dos proporciones, las dos proporciones de la muestra se utilizan para verificar la condición de éxito-fracaso y también para calcular el error estándar, tal como sucedió con una sola proporción.

**6.2.3 Pruebas de hipótesis para -**

Los detalles son muy similares a los de los intervalos de confianza. Sin embargo, esta vez usamos una proporción especial llamada proporción combinada para verificar la condición de éxito y fracaso:



*Utilice la estimación de la proporción combinada cuando es - = 0*

Cuando la hipótesis nula es que las proporciones son iguales, use la proporción combinada () para verificar la condición de éxito-fracaso y estime el error estándar:

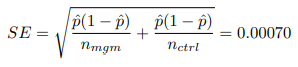


Aquí representa el número de éxitos en la muestra 1 desde:



De manera similar, representa el número de éxitos en la muestra 2.

A continuación, el error estándar se calcula utilizando la proporción combinada, :



*6.2.4 Más sobre las pruebas de hipótesis de 2 proporciones (tema especial)*

Cuando realizamos una prueba de hipótesis de 2 proporciones, generalmente es - = 0. Sin embargo, hay situaciones raras en las que queremos verificar alguna diferencia en y que sea un valor distinto de 0. Por ejemplo, tal vez nos preocupamos por verificar una hipótesis nula donde - = 0.1.11 En contextos como estos, generalmente usamos y para verificar la condición de éxito-fracaso y construir el error estándar.

***6.3 Prueba de bondad de ajuste usando chi-cuadrado (tema especial)***

En esta sección, desarrollamos un método para evaluar un modelo nulo cuando los datos están agrupados. Esta técnica se usa comúnmente en dos circunstancias:

• Dada una muestra de casos que se puede clasificar en varios grupos, determine si la muestra es representativa de la población general.

• Evaluar si los datos se parecen a una distribución particular, como una distribución normal o una distribución geométrica.

Cada uno de estos escenarios se puede abordar utilizando la misma prueba estadística: una prueba de chi-cuadrado.

PAG 286 – 302

En estos problemas, nos gustaría examinar todos los contenedores simultáneamente, no simplemente comparar uno o dos contenedores a la vez, lo que requerirá que desarrollemos una nueva estadística de prueba.

**6.3.1 Creación de una estadística de prueba para tablas unidireccionales**

Cuando se tienen datos estimados en una tabla y proporciones correspondientes a lo esperado, se multiplican las proporciones con el total de los datos estimados para obtener los datos esperados, comparándolos con sus respectivos datos estimados.

La evidencia sólida para la hipótesis alternativa vendría en forma de desviaciones inusualmente grandes en los grupos de lo que se esperaría basándose únicamente en la variación del muestreo.

**6.3.2 El estadístico de la prueba de chi-cuadrado**

En las pruebas de hipótesis anteriores, construimos una estadística de prueba de la siguiente forma:



Esta construcción se basó en (1) identificar la diferencia entre una estimación puntual y un valor esperado si la hipótesis nula era verdadera, y (2) estandarizar esa diferencia utilizando el error estándar de la estimación puntual. Estas dos ideas ayudarán en la construcción de una estadística de prueba apropiada para los datos de recuento. Nuestra estrategia será calcular primero la diferencia entre los recuentos observados y los recuentos que esperaríamos si la hipótesis nula fuera cierta, luego estandarizaremos la diferencia:



El error estándar para la estimación puntual del recuento en los datos agrupados es la raíz cuadrada del recuento bajo el nulo.

La fracción es muy similar a las estadísticas de prueba anteriores: primero calcule una diferencia y luego estandarícela. Estos cálculos también deben completarse para los grupos. Luego, se suman los valores al cuadrado obteniendo el **chi-duadrado**.

El estadístico de prueba , que es la suma de los valores , se utiliza generalmente por estas razones. Estadístico de prueba de chi-cuadrado de . También podemos escribir una ecuación para usando los conteos observados y los conteos nulos:

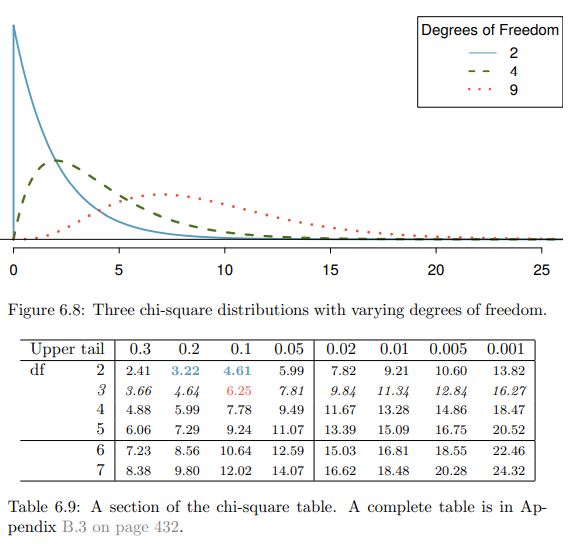


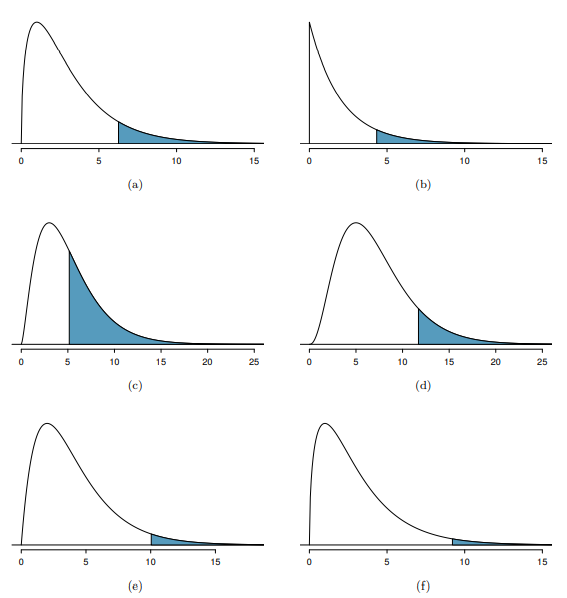
El número final resume la intensidad con la que los recuentos observados tienden a desviarse de los recuentos nulos. En la sección 6.3.4, veremos que si la hipótesis nula es verdadera, entonces sigue una nueva distribución llamada *distribución chi-cuadrado*. Usando esta distribución, podremos obtener un valor p para evaluar las hipótesis.

**6.3.3 La distribución de chi-cuadrado y áreas de búsqueda**

La **distribución de chi-cuadrado** se utiliza a veces para caracterizar conjuntos de datos y estadísticas que siempre son positivas y, por lo general, están sesgadas a la derecha. Recuerde que la distribución normal tenía dos parámetros, la media y la desviación estándar, que podrían usarse para describir sus características exactas. La distribución chi-cuadrado tiene solo un parámetro llamado **grados de libertad (df)**, que influye en la forma, el centro y la extensión de la distribución.

Nuestro principal interés en la distribución chi-cuadrado es el cálculo de valores p, que (como hemos visto antes) está relacionado con encontrar el área relevante en la cola de una distribución. Para hacerlo, se necesita una nueva tabla: la **tabla de chi-cuadrado**, que se muestra parcialmente en la tabla 6.9. Esta tabla es muy similar a la tabla t: examinamos una fila particular para distribuciones con diferentes grados de libertad e identificamos un rango para el área. Una diferencia importante de la tabla t es que la tabla de chi-cuadrado solo proporciona valores de cola superiores.





*Figure 6.10: (a) Distribución chi-cuadrado con 3 grados de libertad, área superior a 6.25 sombreada. (b) 2 grados de libertad, área por encima de 4.3 sombreada. (c) 5 grados de libertad, área por encima de 5.1 sombreada. (d) 7 grados de libertad, área por encima de 11,7 sombreada. (e) 4 grados de libertad, área por encima de 10 sombreada. (f) 3 grados de libertad, área por encima de 9.21 sombreada.*

**6.3.4 Encontrar un valor p para una distribución de chi-cuadrado**

*Prueba de chi-cuadrado para tabla unidireccional*

Supongamos que vamos a evaluar si hay evidencia convincente de que un conjunto de recuentos observados , , ..., en k categorías son inusualmente diferentes de lo que podría esperarse bajo una hipótesis nula. Llame a los recuentos esperados que se basan en la hipótesis nula , , ..., . Si cada recuento esperado es al menos 5 y la hipótesis nula es verdadera, entonces el estadístico de prueba a continuación sigue una distribución de chi-cuadrado con k - 1 grados de libertad:



El valor p para este estadístico de prueba se encuentra observando la cola superior de esta distribución chi cuadrado. Consideramos la cola superior porque valores más grandes de χ2 proporcionarían mayor evidencia en contra de la hipótesis nula.

*CONSEJO: Condiciones para la prueba de chi-cuadrado*

Hay dos condiciones que se deben verificar antes de realizar una prueba de chi-cuadrado:

**Independencia**. Cada caso que aporta un recuento a la tabla debe ser independiente de todos los demás casos de la tabla.

**Tamaño / distribución de la muestra**. Cada escenario particular (es decir, recuento de células) debe tener al menos 5 casos esperados.

No verificar las condiciones puede afectar las tasas de error de la prueba.

Cuando examine una mesa con solo dos contenedores, elija un solo contenedor y use los métodos de una proporción presentados en la Sección 6.1.

***6.4 Prueba de independencia en tablas bidireccionales (tema especial)***

Casos en que también se necesita una variable de resultado para una variable explicativa, comprobando si hay distinción entre ellas o no.

*¿Qué es tan diferente entre las tablas unidireccionales y las tablas bidireccionales?*

Una tabla unidireccional describe los recuentos de cada resultado en una sola variable. Una tabla de dos factores describe los recuentos de combinaciones de resultados para dos variables. Cuando consideramos una tabla de dos factores, a menudo nos gustaría saber si estas variables están relacionadas de alguna manera. Es decir, ¿son dependientes (versus independientes)?

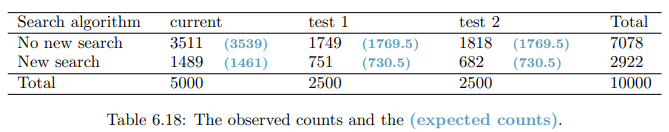
**6.4.1 Recuentos esperados en tablas bidireccionales**

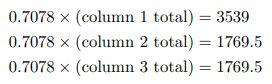
*Ejemplo 6.35* A partir del experimento, estimamos la proporción de usuarios satisfechos con su búsqueda inicial (sin búsqueda nueva) como 7078/10000 = 0,7078. Si realmente no hay diferencia entre los algoritmos y el 70,78% de las personas están satisfechas con los resultados de la búsqueda, ¿cuántas de las 5000 personas en el grupo del "algoritmo actual" se esperaría que no realizaran una nueva búsqueda?

Aproximadamente el 70,78% de los 5000 estarían satisfechos con la búsqueda inicial:

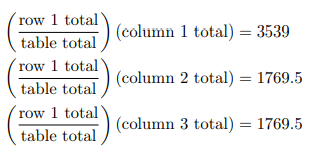
*0.7078 × 5000 = 3539 usuarios*

Es decir, si no hubiera diferencia entre los tres grupos, esperaríamos que 3539 de los usuarios del algoritmo actual no realizaran una nueva búsqueda.





Mirando hacia atrás para ver cómo se calculó la fracción 0.7078, como la fracción de usuarios que no realizaron una nueva búsqueda (7078/10000), estos tres recuentos esperados podrían haberse calculado como:



Esto nos lleva a una fórmula general para calcular los recuentos esperados en una tabla de dos factores cuando nos gustaría probar si existe una fuerte evidencia de una asociación entre la variable de columna y la variable de fila.

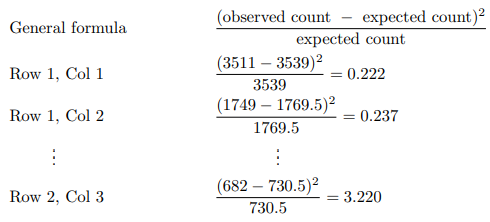
*Calcular los recuentos esperados en una tabla bidireccional*

Para identificar el recuento esperado para la i-ésima fila y la j-ésima columna, calcule:



**6.4.2 La prueba de chi-cuadrado para tablas de dos factores**

El estadístico de prueba de chi-cuadrado para una tabla de dos factores se encuentra de la misma manera que se encuentra para una tabla de un factor. Para cada recuento de tablas, calcule:



Al sumar el valor calculado para cada celda se obtiene el estadístico de prueba de chi-cuadrado :



Al igual que antes, esta estadística de prueba sigue una distribución de chi-cuadrado. Sin embargo, los grados de libertad se calculan de forma un poco diferente para una tabla de dos factores.22 Para las tablas de dos factores, los grados de libertad son iguales a:

*df = (número de filas menos 1) × (número de columnas menos 1)*

*Calcular grados de libertad para una tabla bidireccional*

Al aplicar la prueba de chi-cuadrado a una tabla de dos factores, usamos:



Donde R es el número de filas en la tabla y C es el número de columnas.

***6.5 Prueba de hipótesis de muestra pequeña para una proporción (tema especial)***

En esta sección desarrollamos métodos inferenciales para una sola proporción que son apropiados cuando el tamaño de la muestra es demasiado pequeño para aplicar el modelo normal a . Al igual que los métodos relacionados con la distribución t, estos métodos también se pueden aplicar a muestras grandes.