Estadística No Paramétrica

Clase 10: Test McNemar para muestras relacionadas

Joaquin Cavieres G.

Ingeniería en Estadística

Facultad de Ciencias, Universidad de Valparaíso

Test para muestras relacionadas

El contenido de esta parte del curso se centrará sobre muestras aleatorias relacionadas.

Test para muestras relacionadas

Considere el siguiente problema:

Existen dos mediciones para medir la aprobación de un presidente: aprueba o desaprueba en dos periodos de tiempo determinados (tiempo 1 y tiempo 2). Si consideramos una medición sobre 1600 votantes entonces deberíamos pensar que estas dos variables son relacionadas ya que se les hace la misma pregunta en dos periodos de tiempo distintos ya que sus opiniones deberían ser mas menos similares entre estos dos periodos.

Test para muestras relacionadas

Entonces:

Ya que las observaciones medidas estan agrupadas preferencialmente en la diagonal (794 y 570), nosotros deberíamos rechazar H_0 si consideramos un test de independencia $\chi^2 = 788, df = 1$. Esto es, las mismas personas que aprueban en el primer periodo y siguen apoyando en el segundo periodo de medición (y lo mismo para las otras personas que no apoyan al presidente)

Test para muestras relacionadas

Por ejemplo podemos hacer:

Comparar las proporciones marginales: aprobación en el momento 1, $944/1600 = 0.59$ y aprobación en el momento 2, $880/1600 = 0.55$. La diferencia sería: $0.59 - 0.55 = 0.04$.

Test de Homogeneidad Marginal

Por lo tanto:

Probabilidad de aprueba en el tiempo 1:

$$\mathbb{P}(\text{Aprueba}_{t1}) = \pi_{1+} = \pi_{11} + \pi_{12}$$

Probabilidad de aprueba en el tiempo 2:

$$\mathbb{P}(\text{Aprueba}_{t2}) = \pi_{+1} = \pi_{11} + \pi_{21}$$

Entonces planteamos nuestra hipótesis nula como:

$$H_0 : \pi_{1+} = \pi_{+1}$$

que es equivalente a a proponer una hipótesis que las diagonales son iguales:

$$H_0 : \pi_{12} = \pi_{21}$$

Test de Homogeneidad Marginal

Esta hipótesis:

$$H_0 : \pi_{12} = \pi_{21},$$

también se conoce como hipótesis de simetría y los elementos diagonales no son importantes para este caso particular ya que corresponden a las proporciones de encuestados cuya aprobación / desaprobación se ha mantenido igual a lo largo del tiempo. Toda la información necesaria para determinar si la aprobación ha cambiado o no, está contenida en los elementos fuera de la diagonal.

Test de McNemar

- El Test de McNemar compara las proporciones de dos muestras dicotómicas relacionadas (correlacionadas). Estas dos muestras (variables) pueden ser dos respuestas en un solo individuo o dos respuestas pareadas.
- El test esta basado en una tabla de contingencia de 2×2.
- Las muestras dependientes pueden ocurrir en estudios de pares emparejados y también pueden ocurrir cuando el mismo sujeto se mide en dos momentos diferentes.

Test de McNemar

La idea principal de este test es es verificar si la probabilidad de un evento que se cumple (verdadero) para una determinada variable es igual en los dos niveles de otra variable.

Es un test de homogeneidad marginal (y simetría) para una tabla de 2×2.

Test de McNemar

Supuestos

- Trabajamos con datos pareados
- Se consideran dos variables de tipo binomial y en tablas de contingencia de 2×2
- La suma de eventos que pasan de positivo a negativo y de negativo a positivo debe de ser > 25, de lo contrario, se emplea un test binomial en el que el número de aciertos es el número de eventos que han pasado de positivos a negativos y el número total de intentos es la suma de todos los que han cambiado (de positivos a negativos y de negativos a positivos).

Test de McNemar

Ejemplo

Suponga que a un grupo de sujetos se les aplica un test cuyo resultado sólo contiene dos posibles resultados (positivo y negativo) antes y después de un tratamiento. El objetivo del test es determinar si, mediante el tratamiento, es posible cambiar los resultados de positivo a negativo y lo mismo para uno negativo a positivo.

Test de McNemar

	Despúes positivo	Despúes negativo	Total
Antes positivo	a	b	a+b
Antes negativo	c	d	c+d
Total	a + c	b + d	n=a+b+c+d

Para este caso esperaríamos que las proporciones de pasar negativo a positivo fuesen iguales a las de pasar de positivo a negativo. Esto significa que:

$$\begin{aligned} pc + pd &= pb + pd \\ pc + pd &= pb + pd \end{aligned}$$

Esto nos lleva a proponer que pb = pc, lo que vendría siendo nuestro H_0.

Test de McNemar

Contraste de hipótesis

$$
\begin{aligned}
H_0 : p_b &= p_c \\
H_1 : p_b &\neq p_c
\end{aligned}
$$

Estadístico

El estadístico del test de McNemar sigue una distribución χ^2 con 1 grado de libertad.

$$\chi^2 = \frac{(b-(b+c)/2)^2}{(b+c)/2} + \frac{(c-(b+c)/2)^2}{(b+c)/2} = \frac{(b-c)^2}{b+c}$$

Si b + c < 25 entonces aplicamos:

```
binom.test(x= c, n = c+b, p= 0.5)
```

Test de McNemar

Efecto del tamaño

Ya que queremos determinar si existe relación entre las muestras, el efecto del tamaño también se le conoce como fuerza de asociación.

$$d = \pi_{1+} - \pi_{+1} = \pi_{12} - \pi_{21}$$

y para muestras con un n grande:

$$\hat{d} = \frac{n_{12}}{n} - \frac{n_{21}}{n}$$

La clasificación del efecto esta dada por:

- Pequeño: 0.1
- Mediano: 0.3
- Grande: 0.5

Test de McNemar

Como $\hat{d}$ es insesgado entonces tiene una aproximación Normal con varianza:

$$\begin{aligned} V(\hat{d}) &= n^{-2}V(n_{12} - n_{21}) \\ &= n^{-2}[V(n_{12}) + V(n_{21}) - 2Cov(n_{12}, n_{21})] \\ &= n^{-1}[\pi_{12}(1 - \pi_{12}) + \pi_{21}(1 - \pi_{21}) + 2\pi_{12}\pi_{21}] \end{aligned}$$

Un estimador de la varianza se puede obtener mediante:

$$\hat{V}(\hat{d}) = n^{-1}\left[\frac{n_{12}}{n}(1 - \frac{n_{12}}{n}) + \frac{n_{21}}{n}(1 - \frac{n_{21}}{n}) + 2\frac{n_{12}n_{21}}{n^2}\right]$$

y un intervalo de confianza del 95% como:

$$\hat{d} \pm 1.96\sqrt{\hat{V}(\hat{d})}$$

Test de McNemar

Ahora, suponga que el número total fuera de la diagonal es fijo:

$$n^* = n_{12} + n_{21}$$

Considerando H_0 (mismas proporciones) entonces tendríamos las mismas proporciones de n_{12} y n_{21}, por lo tanto, $n_{12} \sim \text{Bin}(n^*, 0.5)$ (que sería lo mismo para n_{21}. Si n^* es suficientemente grande entonces el test estadístico es:

$$z = \frac{\frac{n_{12}}{n^*} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{n^*}}} = \frac{n_{12} - n_{21}}{\sqrt{n_{12} + n_{21}}}$$

donde $0.5n^* =$ núm esperado de n_{12} bajo H_0. Por tanto la proporción esta determinada por n_{12}/n y la varianza es $\frac{0.5(1-0.5)}{n^*}$ bajo H_0.

Test de McNemar

Bajo H_0, z es asumido $\sim N(0,1)$ y para un $n^* \geq 10$. Alternativamente podemos comparar:

$$z^2 = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}}$$

para χ^2. Este test es válido bajo muestreo multinomial general cuando n^* no es fijo, pero n lo es.

Test de McNemar

$$z^2 = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}}$$

para χ^2. Este test es válido bajo muestreo multinomial general cuando n^* no es fijo, pero n lo es.

Test de McNemar

Ejemplo

Queremos comprobar si un tratamiento de hiervas naturales es capaz de hacer que las personas contesten "Sí" ante solicitudes específicas. Para esto se recoge una muestra de un n número de personas seleccionadas al azar y se les hace una solicitud de ayuda en particular. Las respuestas deben ser Si/No.

Ver respuesta en `R`

Test de McNemar

Ejemplo 2

Mediante un estudio se quiere determinar si la intención de voto se ve afectada por el COVID en la población. Para ello, antes de la votación, se pregunta a las 1600 personas del grupo si irían a votar. Tras enterarse que el COVID ha aumentado levemente (n casos) entonces se les consulta de nuevo a las mismas personas. ¿Existen evidencias significativas de que la decisión de ir a votar a cambiado debido a la enfermedad?

Ver respuesta en `R`