



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO
ESCUELA SUPERIOR DE CIUDAD SAHAGÚN

Pruebas de bondad de ajuste

Área Académica: Licenciatura en Ingeniería Industrial

Profesor(a): Mtra. Ma. Guadalupe Vera Correa

Periodo: Julio - diciembre 2017

Pruebas de bondad de ajuste

RESUMEN

En esta sección se estudia el caso en que cada elemento de una población corresponde a una y sólo a una de varias clases o categorías. A estas poblaciones se les conoce como **poblaciones multinomiales**. La distribución multinomial se puede entender como una extensión de la distribución binomial al caso en el que hay tres o más categorías de resultados. En cada ensayo de un experimento multinomial uno y sólo uno de los resultados ocurre. Se supone que cada ensayo del experimento es independiente y que en todos los ensayos las probabilidades para los resultados permanecen constantes.

ABSTRACT

In this section we study the case where each element of a population corresponds to one and only to one of several classes or categories. These populations are known as populations multinomiales. The multinomial distribution can be understood as an extension of the distribution Binomial to the case where there are three or more categories of results. In each trial of a multinomial experiment one and only one of the results occurs. It is assumed that each trial of the experiment is independent and that in all trials the probabilities for the results remain constant.

Keywords: Multinomial, Hypothesis testing, Null hypothesis, Alternative hypothesis, Statistical test



PRUEBAS DE BONDAD DE AJUSTE

Estas pruebas permiten verificar que la población de la cual proviene una muestra tiene una distribución especificada o supuesta.

Sea X : variable aleatoria poblacional

$f_0(x)$ la distribución (o densidad) de probabilidad especificada o supuesta para X

Se desea probar la hipótesis: $H_0: f(x) = f_0(x)$

En contraste con la hipótesis alterna:

$H_a: f(x) \neq f_0(x)$ (negación de H_0)



PRUEBA JI-CUADRADO

Esta prueba es aplicable para variables aleatorias discretas o continuas. Sea una muestra aleatoria de tamaño n tomada de una población con una distribución especificada $f_0(\mathbf{x})$ que es de interés verificar.

Suponer que las observaciones de la muestra están agrupadas en k clases, siendo o_i la cantidad de observaciones en cada clase $i = 1, 2, \dots, k$

Con el modelo especificado $f_0(\mathbf{x})$ se puede calcular la probabilidad p_i que un dato cualquiera pertenezca a una clase i .



Con este valor de probabilidad se puede encontrar la frecuencia esperada e_i para la clase i , es decir, la cantidad de datos que según el modelo especificado deberían estar incluidos en la clase i : $e_i = p_i n$, $i = 1, 2, \dots, k$

Tenemos entonces dos valores de frecuencia para cada clase i

- f_i : frecuencia observada (corresponde a los datos de la muestra)
- e_i : frecuencia esperada (corresponde al modelo propuesto)



La teoría estadística demuestra que la siguiente variable es apropiada para realizar una prueba de bondad de ajuste:

Estadístico para la prueba de bondad de ajuste Ji-cuadrado

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - e_i)^2}{e_i}$$



distribución Ji-cuadrado con $v=k-r-1$ grados de libertad

Dado un nivel de significancia α se define un valor crítico χ^2_{α} para el rechazo de la hipótesis propuesta

$$H_0: f(x) = f_0(x).$$



- En esta sección se estudia el caso en que cada elemento de una población corresponde a una y sólo a una de varias clases o categorías. A estas poblaciones se les conoce como **poblaciones multinomiales**.

- Prueba de bondad de ajuste

El propósito de la prueba es averiguar si existen diferencias estadísticamente significativas entre la distribución observada y la distribución esperada.

En la prueba se plantean las siguientes hipótesis estadísticas:

Hipotesis nula

H_0

Hipotesis alternativa

H_a



Ejemplo:

Considere un estudio sobre participación en el mercado realizado por la empresa Scott Marketing Research. A lo largo de los años las participaciones en el mercado se han estabilizado en 30% para la empresa A, 50% para la empresa B y 20% para la empresa C. Recién la empresa C ha elaborado un nuevo y mejorado producto para sustituir a uno de sus productos en el mercado y pidió a la empresa Scott Marketing Research que determinara si el nuevo producto modificaría su participación en el mercado.



Solución

En este caso, la población de interés es multinomial, cada cliente se clasifica como cliente de la empresa A, de la empresa B o de la empresa C. De manera que se tiene una población multinomial con tres resultados. Para las proporciones se usa la notación siguiente

- p_A = participación en el mercado de la empresa A
- p_B = participación en el mercado de la empresa B
- p_C = participación en el mercado de la empresa C



Scott Marketing Research realizará un estudio muestral y calculará la proporción que prefiere el producto de cada empresa. Después aplicará una prueba de hipótesis para ver si el nuevo producto modifica las participaciones en el mercado. Suponga que el nuevo producto de la empresa C no modifica las participaciones en el mercado; entonces, las hipótesis nula y alternativa serán las siguientes.

- $H_0: p_A = 0.30, p_B = 0.50, \text{ y } p_C = 0.20$
- H_a : Las proporciones poblacionales no son $p_A = 0.30, p_B = 0.50, \text{ y } p_C = 0.20$



Si los resultados muestrales llevan al rechazo de H_0 Scott Marketing Research tendrá evidencias de que la introducción del nuevo producto afecta las participaciones del mercado.

Considere que para este estudio la empresa de investigación de mercado ha empleado un panel de 200 consumidores. A cada individuo se le pide que indique su preferencia entre el producto de la empresa A, el producto de la empresa B o el nuevo producto de la empresa C. Las 200 respuestas obtenidas se presentan a continuación en forma resumida.

Frecuencia esperada

Producto de la Empresa A	Producto de la Empresa B	Producto de la Empresa C
48	98	54



Ahora se realiza la **prueba de bondad de ajuste** para determinar si la muestra de las 200 preferencias de los clientes coincide con la hipótesis nula.

La prueba de bondad de ajuste se basa en la comparación de los resultados muestrales *observados* con los resultados *esperados*, bajo la suposición de que la hipótesis nula es verdadera.

Por tanto, el paso siguiente es calcular las preferencias esperadas en los 200 clientes, con el supuesto de que $p_A = 0.30$, $p_B = 0.50$ y $p_C = 0.20$ hacerlo dará los resultados esperados.



Frecuencia observada

Producto de la Empresa A	Producto de la Empresa B	Producto de la Empresa C
$200(0.30) = 60$	$200(0.50) = 100$	$200(0.20) = 40$

Como se observa, la frecuencia esperada de cada categoría se encuentra multiplicando el tamaño de la muestra, 200, por la proporción hipotética de esa categoría.



- * En la prueba de bondad de ajuste lo que interesa son las diferencias entre frecuencias observadas y frecuencias esperadas. Grandes diferencias entre frecuencias observadas y frecuencias esperadas harán dudar sobre la exactitud de las proporciones o participaciones en el mercado hipotéticas.
- * El que las diferencias entre frecuencias observadas y esperadas sean “grandes” o “pequeñas” es una cuestión que se determina con ayuda del estadístico de prueba.



ESTADISTICO DE PRUEBA PARA LA PRUEBA DE BONDAD DE AJUSTE

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - e_i)^2}{e_i}$$

Donde:

f_i = frecuencia observada en la categoría i

e_i = frecuencia esperada en la categoría i

k = numero de categorías



Ahora, de regreso con Scott Marketing Research, los datos muestrales se emplearán para probar la hipótesis de que en la población multinomial las proporciones sigan siendo $p_A = 0.30$, $p_B = 0.50$ y $p_C = 0.20$.

El nivel de significancia que se va a usar es 0.05. Mediante las frecuencias observadas y esperadas se calcula el valor del estadístico de prueba utilizando la formula anterior.



Categoría	Proporción hipotética	Frecuencia observada f_i	Frecuencia esperada e_i	Diferencia $(f_i - e_i)$	Cuadrado de la diferencia $(f_i - e_i)^2$	$(f_i - e_i)^2 / e_i$
Empresa A	0.30	48	60	48-60=-12	$(-12)^2=144$	144/60=2.40
Empresa B	0.50	98	100	98-100=-2	$(-2)^2=4$	4/100=0.04
Empresa C	0.20	54	40	54-40=14	$(14)^2=196$	196/40=4.90
Total		200				$\chi^2 = 7.34$

Tabla a



Como las frecuencias esperadas son todas 5 o más, se calcula el estadístico de prueba chi-cuadrada como se muestra en la **tabla a**. Se obtiene $\chi^2 = 7.34$

La hipótesis nula se rechaza si las diferencias entre las frecuencias observadas y esperadas son *grandes*. Diferencias grandes entre las frecuencias esperadas y observadas darán un valor grande del estadístico de prueba. Entonces, la prueba de bondad de ajuste siempre será una prueba de la cola superior.



El área en la cola superior se emplea en el método del estadístico de prueba y en el método del valor- p para determinar si se puede rechazar la hipótesis nula. Para $k-1 = 3-1= 2$ grados de libertad.

En la tabla de la distribución chi-cuadrada se observan los datos siguientes

Área en la cola Superior	0.10	0.05	0.025	0.01	0.005
Valor $\chi^2 = (2gl)$	4.605	5.991	7.378	9.210	10.597

$\chi^2=0.734$



El estadístico de prueba $X^2 = 7.34$ se encuentra entre 5.991 y 7.378. Por consiguiente, el área correspondiente en la cola superior o valor- p debe estar entre 0.05 y 0.025. Como el valor- $p \leq \alpha = 0.05$, se rechaza H_0 y se concluye que la introducción del nuevo producto de la empresa C sí modifica la estructura de la participación de mercado.



En lugar del método del valor- p se puede utilizar el método del valor crítico con el que se llega a la misma conclusión.

Como $\alpha = 0.05$ y los grados de libertad son 2, el valor crítico para el estadístico de prueba es $\chi^2_{0.05} = 5.991$. La regla de rechazo de la cola superior es:

Rechazar H_0 si $X^2 \geq 5.991$



Referencias

Anderson, Williams, Learning. *Estadística para Administración y Economía*. Decima Edición.

