

Prueba de bondad de ajuste

Jessica Nathaly Pulzara Mora
jessica.pulzara@udea.edu.co

Departamento de ingeniería de sistemas



**UNIVERSIDAD
DE ANTIOQUIA**

Prueba de bondad de ajuste

Una prueba de Bondad de Ajuste es un procedimiento especial para determinar si un conjunto de datos proviene de cierta distribución conocida.

- Las pruebas de hipótesis que discutimos anteriormente están diseñadas para resolverse cuando la distribución de la población es conocida, y sus parámetros hacen parte de la hipótesis.
- Otro tipo de pruebas de hipótesis también son importantes: cuando no conocemos la distribución de la población, y deseamos probar que cierta distribución es un modelo satisfactorio.

Pruebas de bondad de ajuste

Existen varias pruebas de bondad de ajuste:

- Prueba Chi-cuadrado: la más utilizada cuando los datos son discretos.
- Prueba Chi-cuadrado para distribuciones continuas.
- Otras pruebas para variables continuas que no hacen parte del curso: Kolmogorov - Smirnov, Shapiro - Wilks, Cramer Von Mises, Jarque Bera, entre otras.

Prueba Chi-cuadrado

Prueba Chi-cuadrado

Se dispone de n observaciones independientes de una v.a X , estas observaciones son agrupadas en k posibles categorías o clases (por ejemplo, al lanzar un dado se tienen 6 clases: 1,2,3,4,5,6).

Sea p_i la probabilidad de clasificar en la categoría i .

Prueba Chi-cuadrado

Las hipótesis que se desean poner a prueba son:

$$H_0 : F(x) = F_0(x)$$

$$H_1 : F_0(x) \text{ no es la f.d.p asociada a la muestra}$$

O alternativamente:

$$H_0 : p_i = p_{i0} \quad i = 1, 2, \dots, k$$

$$H_1 : \exists_j \text{ tal que } p_j \neq p_{j0}$$

Si F_0 esta claramente especificada, es posible conocer los valores particulares para los p_i

Prueba Chi-cuadrado datos discretos

Se tiene lo siguiente:

$$H_0 : p_i = p_{i0} \quad i = 1, 2, \dots, k$$

$$H_1 : \exists j \text{ tal que } p_j \neq p_{j0}$$

Categoría/v.a	Frec. Observada	Probabilidad	Frec. Esperada
1	n_1	p_1	np_1
2	n_2	p_2	np_2
\vdots	\vdots	\vdots	\vdots
i	n_i	p_i	np_i
\vdots	\vdots	\vdots	\vdots
k	n_k	p_k	np_k

- n es el tamaño de la muestra.
- k es el número de categorías.

Prueba Chi-cuadrado

El estadístico de prueba es el siguiente:

$$X_c = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} \sim \chi^2(k - 1)$$

- La región de rechazo es: $\{X_c : X_c > \chi^2_{\alpha}(k - 1)\}$
- El valor $p = P(\chi^2 > X_c)$.
- H_0 se rechaza si valor $p < \alpha$.

Nota: np_i debe ser mayor o igual que 5.

Ejemplo

Para determinar si un dado está cargado o no, éste es lanzado 600 veces y se anota cuantas veces aparece cada cara. Los resultados fueron los siguientes:

Cara	1	2	3	4	5	6
Frec. Observada	85	94	108	112	98	103

¿Qué podemos afirmar sobre el dado?

Cara	1	2	3	4	5	6
Frec. Observada	85	94	108	112	98	103
Frec. Esperada	100	100	100	100	100	100

Hipótesis

$$H_0 : p_i = \frac{1}{6} \quad i = 1, 2, \dots, k$$

$$H_1 : \exists_j \text{ tal que } p_j \neq \frac{1}{6}$$

$X_c = 4.82$, valor $p = 0.1$, por lo tanto, con un 95 % de confianza, no hay evidencia para afirmar que el dado está cargado.