

Tema 7. Kolmogorov-Smirnov y Lilliefors

Dr. David Castilla Espino

Inferencia Estadística

CA UNED de Huelva, "Profesor Dr. José Carlos Vílchez Martín"

Referencias

- Casas, JM, C. Sánchez & P. Cortiñas (2018): Inferencia Estadística para Economía. Editorial Universitaria Ramón Areces, Madrid (Capítulo 7).
- Novales (1997), Ruiz Maya & Martín Pliego (1999),...

Contraste de Kolmogorov-Smirnov

- El contraste de Kolmogorov-Smirnov es un contraste no paramétrico de bondad del ajuste.
- **Idea básica:** Se utiliza para contrastar si un conjunto de datos muestrales pueden considerarse procedentes de una distribución determinada (normal, exponencial...).
- **Hipótesis del contraste:**
 - $H_0: F_n(X) = F_0(X)$ $H_0: F_n(X) \geq F_0(X)$ $H_0: F_n(X) \leq F_0(X)$
 - $H_1: F_n(X) \neq F_0(X)$ $H_1: F_n(X) < F_0(X)$ $H_1: F_n(X) > F_0(X)$
- **Uso:** Es una alternativa a la prueba Chi-cuadrado cuando el modelo propuesto bajo la hipótesis nula es de tipo continuo y el tamaño muestral es pequeño. Este contraste no requiere que las observaciones muestrales se agrupen en intervalos o clases; aunque exige que los parámetros de la distribución teórica sean conocidos.

Contraste de Kolmogorov-Smirnov

- **Lógica:** Habrá evidencias para rechazar la hipótesis nula cuando la discrepancia medida por medio de la diferencia entre $F_n(X)$ y $F_0(X)$ sea lo suficientemente grande.
- **Pasos para resolución:**
 - Cálculo de la $F_n(X) = N(x)/n$.
 - Cálculo de la $F_0(X)$ en los valores de la muestra de datos.
 - Cálculo de $|F_n(X) - F_0(X)|$, $F_0(X) - F_n(X)$ o $F_n(X) - F_0(X)$ según el contraste sea bilateral, unilateral por la izquierda o unilateral por la derecha.
 - Cálculo del estadístico de contraste y determinación de la región crítica conforme a la tabla que se muestra a continuación:

H_0	H_1	Estadístico	Región crítica
$F_n(X)=F_0(X)$	$F_n(X)\neq F_0(X)$	$D_n=\max F_n(X)-F_0(X) $	$D_n>D_\alpha$ donde $P(D_n>D_\alpha/H_0)=\alpha$
$F_n(X)\geq F_0(X)$	$F_n(X)<F_0(X)$	$D_n^+=\max[F_0(X)-F_n(X)]$	$D_n^+>D_\alpha$ donde $P(D_n^+>D_\alpha/H_0)=\alpha$
$F_n(X)\leq F_0(X)$	$F_n(X)>F_0(X)$	$D_n^-=\max[F_n(X)-F_0(X)]$	$D_n^->D_\alpha$ donde $P(D_n^->D_\alpha/H_0)=\alpha$

Contraste de normalidad de Lilliefors

- El contraste de normalidad de Lilliefors es un contraste no paramétrico de bondad del ajuste.
- **Idea básica:** Se utiliza para contrastar si un conjunto de datos muestrales pueden considerarse procedentes de una distribución Normal.
- **Hipótesis del contraste:**
 - $H_0: F_n(X) = F_0(X)$ tal que $X \rightarrow N(\mu, \sigma)$.
 - $H_1: F_n(X) \neq F_0(X)$ tal que $X \rightarrow N(\mu, \sigma)$.
- **Uso:** Esta prueba se utiliza cuando el modelo propuesto se caracteriza por que la hipótesis nula Normal, el tamaño muestral es pequeño y los parámetros de la distribución son desconocidos. Este contraste no requiere que las observaciones muestrales se agrupen en intervalos o clases.

Contraste de normalidad de Lilliefors

- **Lógica:** Habrá evidencias para rechazar la hipótesis nula cuando la discrepancia medida por medio de la diferencia entre $F_n(X)$ y $F_0(X)$ sea lo suficientemente grande.
- **Pasos para resolución:**
 - Cálculo de la $F_n(X) = N(x)/n$.
 - Cálculo de la $F_0(X)$ en los valores de la muestra de datos. Se requiere tipificar la variable X si se va a emplear las tablas de la distribución Normal Tipificada.
 - Cálculo de $|F_n(X) - F_0(X)|$.
 - Cálculo del estadístico de contraste y determinación de la región crítica conforme a la tabla que se muestra a continuación:

Contraste de normalidad de Lilliefors			
H_0	H_1	Estadístico	Región crítica
$F_n(X)=F_0(X)$	$F_n(X)\neq F_0(X)$	$D'_n=\max F_n(X)-F_0(X) $	$D'_n>D'_\alpha$ donde $P(D'_n>D'_\alpha/H_0)=\alpha$

Ejemplo

Compruebe si la edad de los habitantes de un pequeño pueblo se distribuye de acuerdo con una familia de distribuciones de probabilidad $U(a=0, b=100)$ y $N(\mu, \sigma)$ empleando los contrastes de hipótesis no paramétricos estudiados. Utilice para ello la muestra aleatoria simple de edades 40 habitantes del pueblo que se muestra a continuación.

5	8	11	13	16	20	20	28	30	36
41	41	44	45	46	53	53	54	56	57
58	60	63	65	71	72	77	85	87	89
90	90	91	92	95	95	97	98	98	100

Nótese que la distribución de probabilidad $U(a=0, b=100)$ tiene la siguiente $F(X)$:

$$F(X) = \begin{cases} 1 & x \geq 100 \\ \frac{x}{100} & x \in (0, 100) \\ 0 & x \leq 0 \end{cases}$$