



PRUEBA JI-
CUADRADA

BONDAD DE AJUSTE

¿Qué es bondad de ajuste?

Como tomadores de decisiones que utilizamos la estadística, necesitaremos escoger cierta distribución de probabilidad para representar la distribución de los datos que tengamos que analizar.

La prueba ji-cuadrada nos permite probar si existe una diferencia significativa entre una distribución de frecuencias observadas y una distribución de frecuencias teórica. De esta manera, podemos determinar la bondad de ajuste a una distribución teórica, es decir, podemos determinar si los datos si debemos creer que los datos observados constituyen una muestra obtenida de la distribución teórica hipotética.

Prueba ji-cuadrada

Hipótesis a contrastar

H0: Los datos observados conforman una muestra de una población con distribución *teórica*

Ha: los datos observados provienen de una población con otra distribución

Estadístico de Prueba

$$X^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

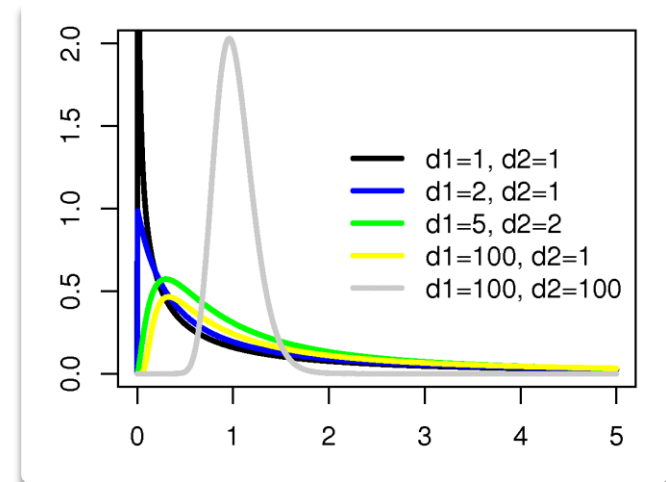
donde f_o es la frecuencia observada y f_e es la frecuencia esperada bajo la distribución teórica.

Rechazando H0 si $X^2 \geq X^2_{\alpha, k-1-t}$ donde k es el número de sumandos al calcular el estadístico de prueba y t es el número de parámetros estimados

Identificación

La identificación de qué distribución probar al considerar un conjunto de datos puede ser mediante información de trabajos previos o mediante el uso de los histogramas comparados con la gráfica generada en la familia de distribuciones teóricas.

Los histogramas son generados comúnmente de las tablas de frecuencias, en caso de variables continuas, o de los conteos generados en variables discretas.



Ejemplo 1

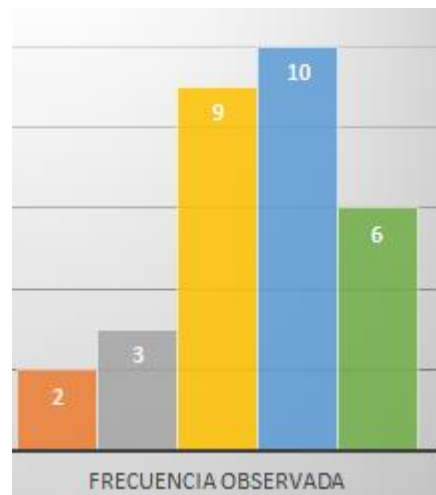
Las siguientes son las calificaciones finales del semestre anterior del grupo 02 en la materia A

70	62	69	61	67	58
67	75	69	67	75	72
66	72	77	74	73	68
73	68	71	64	74	73
72	66	65	72	75	73

Crear la tabla de frecuencias

Numero de clase	Intervalo de Clase		frecuencia observada
	Limite Inferior	Limite superior	
1	58	61.8	2
2	61.8	65.6	3
3	65.6	69.4	9
4	69.4	73.2	10
5	73.2	77	6

Graficar



Ejemplo 1

Hipótesis a contrastar

H0: Los datos observados conforman una muestra de una población con distribución *Normal (aquí nos interesa sólo que sea normal)*

Ha: los datos observados provienen de una población con otra distribución

La media y varianza (según los estimadores revisados en inferencia) son estimados por 70 y 22, respectivamente. De aquí que $t=2$

Numero de clase	Intervalo de Clase		frecuencia observada	probabilidad esperada	frecuencia esperada	cocientes
	Limite Inferior	Limite superior				
1	58	61.8	2	0.04021073	1.20632192	0.5221864
2	61.8	65.6	3	0.13389011	4.01670327	0.25734675
3	65.6	69.4	9	0.27500513	8.25015396	0.06815256
4	69.4	73.2	10	0.30335115	9.10053451	0.08890007
5	73.2	77	6	0.24754288	7.42628634	0.27393136

30

1

30 1.21051714

Ejemplo 1

De la tabla se tiene que el estadístico de prueba esta dado por $X^2 = 1.21$ mientras que el valor de tabla $X^2_{\alpha, k-1-t} = X^2_{0.05, 5-1-2} = 5.9914$. Ya que rechazamos H_0 si $X^2 \geq X^2_{\alpha, k-1-t}$ se tiene 1.21 no es mayor que 5.99, por lo cual no se tiene suficiente evidencia para rechazar H_0 .

Así con 95% de confianza tenemos que no existe evidencia suficiente para decir que los datos observados provienen de una población con otra distribución que no sea la distribución normal.

Ejemplo 2

¿Qué pasa si se quiere verificar bajo una distribución normal específica?

No sólo el valor de t cambia, sino los valores tanto de media y varianza propuestos se incluyen al llenar la tabla para calcular los cocientes a sumar para calcular el estadístico de prueba.

Considere por ejemplo si se tienen las hipótesis a contrastar

H_0 : Los datos observados conforman una muestra de una población con distribución Normal con media 75 y varianza 5

H_a : los datos observados provienen de una población con otra distribución

Ejemplo 2

De la tabla se tiene que el estadístico de prueba esta dado por $X^2 = 74829729.7$ mientras que el valor de tabla $X^2_{\alpha, k-1-t} = X^2_{0.05, 5-1} = 9.4877$. Ya que rechazamos H_0 si $X^2 \geq X^2_{\alpha, k-1-t}$ se tiene que se cumple la desigualdad de la región de rechazo.

Así con 95% de confianza tenemos que no existe evidencia suficiente para decir que los datos observados provienen de una población con distribución normal con media 75 y varianza 5.

Numero de clase	Intervalo de Clase		frecuencia observada	probabilidad esperada	frecuencia esperada	cocientes
	Limite Inferior	Limite superior				
1	58	61.8	2	1.78238E-09	5.3471E-08	74806433.1
2	61.8	65.6	3	1.31213E-05	0.00039364	22857.5727
3	65.6	69.4	9	0.006119908	0.18359723	423.366729
4	69.4	73.2	10	0.20428129	6.12843869	2.44580843
5	73.2	77	6	0.78958568	23.6875704	13.2073548

30

1

30 74829729.7

Ejemplo 3

¿Qué pasa en el caso discreto?

Se lanzó una moneda 30 veces obteniendo un total de 18 caras, ¿se puede decir que la moneda es una moneda balanceada?

H0: Los datos observados conforman una muestra de una población con distribución binomial con probabilidad 0.5, la moneda es balanceada.

Ha: los datos observados provienen de una población con otra distribución

De la tabla se tiene que el estadístico de prueba esta dado por $X^2 = 1.2$ mientras que el valor de tabla

$X^2_{\alpha, k-1-t} = X^2_{0.05, 2-1} = 3.8414$. Ya que rechazamos H0 si $X^2 \geq X^2_{\alpha, k-1-t}$

no se rechaza H0 dado que no se cumple la desigualdad anteriormente mencionada. Así con 95% de confianza la moneda es legal.

resultado	frecuencia observada	probabilidad	frecuencia esperada	cociente
cara	18	0.5	15	0.6
cruz	12	0.5	15	0.6

30

1

30

1.2

Consideraciones finales

- ▶ Se emplean algunas técnicas revisadas en el curso de inferencia como la creación de tabla y las estimaciones según la distribución de interés.
- ▶ Es necesario crear tablas de frecuencias al manejar variables continuas, pero de igual forma se puede emplear la prueba de bondad de ajuste en variables discretas.
- ▶ $\chi^2_{\alpha, k-1-t}$ se calcula mediante el uso de la cola derecha.
- ▶ El valor de t dependerá si la hipótesis alternativa es simple o compuesta.
- ▶ Generalmente k no es igual a n.
- ▶ El uso del histograma es sólo a manera de sugerencia al momento de elegir la distribución a probar, pero recuerde que algunas distribuciones pueden llegar a confundirse, por ejemplo, una distribución t con una distribución z. Usted puede realizar múltiples pruebas.