PRUEBAS DE BONDAD DE AJUSTE

- Las pruebas de bondad de ajuste son pruebas de hipótesis para verificar si los datos observados en una muestra aleatoria se ajustan con algún nivel de significancia a determinada distribución de probabilidad (uniforme, exponencial, normal, poisson, u otra cualquiera).
- La hipótesis nula H_o indica la distribución propuesta, mientras que la hipótesis alternativa H_1, nos indica que la variable en estudio tiene una distribución que no se ajusta a la distribución propuesta.

$$H_o: f(x)=f_o(x)$$

$$H_1: f(x)\neq f_o(x)$$

PRUEBAS DE BONDAD DE AJUSTE

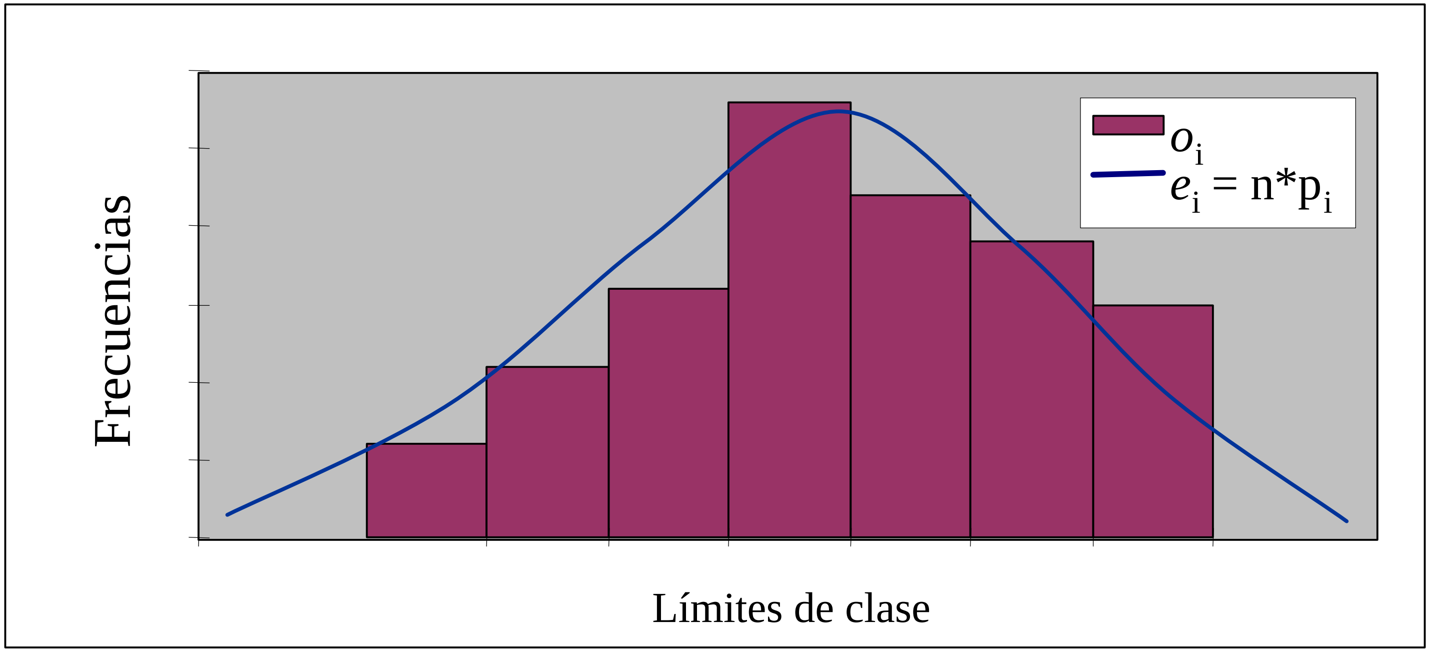
Para realizar la prueba, se clasifican los datos observados en k clases o categorías, y se contabiliza el número de observaciones en cada clase, para posteriormente comparar la frecuencia observada en cada clase con la frecuencia que se esperaría obtener en esa clase si la hipótesis nula es correcta.

k = No. de clases, $k>2$

o_i = Frecuencia observada en la clase i

e_i = Frecuencia esperada en la clase i, si H_o es correcta.

PRUEBAS DE BONDAD DE AJUSTE



Las pruebas de bondad de ajuste comparan la frecuencia observada con la frecuencia esperada en cada clase.

$e_i = n*p_i$, donde:

n=tamaño de la muestra,

p_i=área bajo la curva $f_o(x)$ en el intervalo lim_{sup}-lim_{inf} de la clase i

Si $f_o(x)$ es continua, entonces:

$$p_i = \int_{lím_{\inf i}}^{lím_{\sup i}} f_o(x)dx$$

PRUEBAS DE BONDAD DE AJUSTE
- LA PRUEBA JI-CUADRADA -

Existen varios procedimientos para probar la bondad de ajuste de una distribución a los datos observados en una muestra, uno de ellos es la prueba Ji-cuadrada, que se basa en el estadístico de prueba:

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

El cual tiene distribución Ji-cuadrada con k-r-1 grados de libertad. Si las diferencias o_i-e_i son pequeñas, el valor del estadístico es pequeño, por el contrario si esas diferencias son grandes (lo observado no se ajusta a lo propuesto), el valor del estadístico es grande, por lo tanto, la región de rechazo de la hipótesis nula se ubica en la cola superior de la distribución Ji-cuadrada al nivel de significancia α.

$$RR : \{Y > c^{2}_{a,k-r-1}\}$$

Donde: k = No. de clases.

r = no. de parámetros estimados en $f_o(x)$ para encontrar e_i

RECOMENDACIONES PARA REALIZAR LA PRUEBA

- El tamaño de la muestra deberá ser *moderadamente grande*, pues si la muestra es muy pequeña no se podrá formar un número suficiente de clases y si la muestra es muy grande la prueba conducirá al rechazo casi con seguridad. Se sugiere que n sea aproximadamente igual a 5 veces el número de clases.
- Se recomienda clasificar la muestra en mínimo cinco clases y máximo diez.
- Hacer que toda frecuencia observada o esperada no sea menor que cinco, esto puede lograrse combinando clases vecinas, pero para cada par de clases que se combinan, el número de grados de libertad debe reducirse en uno (k es el número de clases efectivas en la tabla de frecuencias).
- Si $fo(x)$ es continua:
  - Para la primera clase, calcular p_1 considerando el intervalo desde $-\infty$ hasta el límite superior de la clase.
  - Para la última clase, calcular p_k considerando el intervalo desde el límite inferior de la clase hasta $+\infty$.

PRUEBA DE BONDAD DE AJUSTE CHI-CUADRADA

TAREA: REVISAR LOS CÁLCULOS DE ESTE EJEMPLO

Ejercicio 11.47, pag. 404, Hines

El número de unidades defectuosas encontradas cada día por un probador de circuitos en un proceso de ensamble se muestra a continuación.
¿Es razonable concluir que estos datos provienen de una distribución normal?

Ho: f(x)=fo(x); Normal

Ho sin especificar media y varianza

media	Desv. Est.
25.26	9.233465719

Datos estimados a partir de la muestra.

α= 0.05
k= 7
r= 2
ν= 4

Datos Originales

intervalo		frec
0	10	6
11	15	11
16	20	16
21	25	28
26	30	22
31	35	19
36	40	11
41	45	4

* Como la última clase tiene frecuencia menor a 5, se combina con la clase

Tabla de frecuencias de la muestra

límites aparentes		(fronteras) límites reales		xi	Frecuencia observada fi = Oi	Frecuencia relativa f'i
0	10	-0.5	10.5	5	6	0.0513
11	15	10.5	15.5	13	11	0.0940
16	20	15.5	20.5	18	16	0.1368
21	25	20.5	25.5	23	28	0.2393
26	30	25.5	30.5	28	22	0.1880
31	35	30.5	35.5	33	19	0.1624
36	45	35.5	45.5	40.5	15	0.1282
Sumas:				n=	117	1

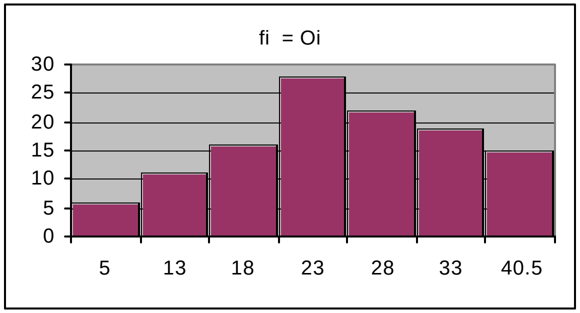


Tabla de frecuencias esperadas si Ho es correcta

intervalo de Z bajo Ho Lím$_{inf}$	Lím$_{sup}$		p_i	esperada bajo Ho $e_i = n*p_i$	$(O_i-e_i)^2/e_i$
$-\infty$	-1.60	P(Z<-1.6) =	0.0550	6.43	0.02870594
-1.60	-1.06	P(-1.6<Z<-1.06) =	0.0903	10.56	0.0181057
-1.06	-0.52	P(-1.06<Z<-0.52) =	0.1578	18.47	0.32954852
-0.52	0.03	P(-0.52<Z<0.03) =	0.2073	24.25	0.57975963
0.03	0.57	P(0.03<Z<0.57) =	0.2044	23.92	0.15419785
0.57	1.11	P(0.57<Z<1.11) =	0.1515	17.72	0.09193601
1.11	$+\infty$	P(Z>1.11) =	0.1337	15.65	0.02669689
			1.0000	χ^2_m=	1.229

Rechazar Ho si $\chi^2_m > \chi^2_{\alpha, k-1-r} =$ 9.48772846

Conclusión: No se rechaza Ho

Valor P = 0.87331

Ho: f(x)=Uniforme

PRUEBA DE BONDAD DE AJUSTE CHI-CUADRADA

Ejercicio 10.5, pag. 375, Canavos

Una organización de seguridad vial desea determinar si el número de accidentes fatales se encuentra distribuido de igual forma para el color de los automóviles involucrados.
De una muestra aleatoria de 600 accidentes se obtuvieron los siguientes resultados.
¿Existe alguna razón para creer que las proporciones de color no son idénticas?

α= 0.01
k= 6
r= 1
ν= 4

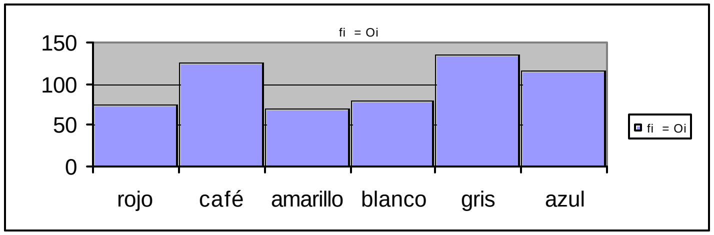
clase	frecuencia observada fi = Oi	frecuencia relativa f'i
rojo	75	0.1250
café	125	0.2083
amarillo	70	0.1167
blanco	80	0.1333
gris	135	0.2250
azul	115	0.1917

n= 600

pi	frecuencia esperada bajo Ho ei = n*pi	$(Oi-ei)^2/ei$
0.1667	100.00	6.25
0.1667	100.00	6.25
0.1667	100.00	9
0.1667	100.00	4
0.1667	100.00	12.25
0.1667	100.00	2.25

1.0000

c^{2}_{m}= 40.000



Rechazar Ho si

c^{2}_{m} > $c^{2}_{\alpha,\, k-1-r}$ = 13.2766986

Se rechaza Ho

Valor P = 0.00000

PRUEBA DE BONDAD DE AJUSTE CHI-CUADRADA

Ho: F(x)=Ver enunciado

Ejercicio 10.1, pag. 374, Canavos

Según lo registros de una tienda modas para esta temporada, de los vestidos el 50% se venden a precio normal, el 25% con un descuento de 20%, el 15% con un descuento de 40% y los restantes con una reducción en su precio de 60%.

Para esta temporada se adquirieron 300 vestidos y su venta fue de la siguiente forma:
¿Existe alguna razón para creer que la disminución en ventas fue diferente en esta temporada?

$\alpha = 0.05$
$k = 4$
$r = 0$
$\nu = 3$

venta real:

clase	frecuencia observada $f_i = O_i$	frecuencia relativa f'_i
Precio Normal	140	0.4667
20% desc.	90	0.3000
40% desc.	30	0.1000
60% desc.	40	0.1333

n= 300

pi	frecuencia esperada bajo Ho $e_i = n \cdot p_i$	$(O_i-e_i)^2/e_i$
0.5000	150.00	0.66666667
0.2500	75.00	3
0.1500	45.00	5
0.1000	30.00	3.33333333

1.0000

$c^2_m =$ 12.000

Rechazar Ho si

$c^2_m > c^2_{\alpha,\, k-1-r} =$ 7.8147247

Se rechaza Ho

Valor P = 0.00738