

## Problemas resueltos. Tema 12

1.- En una partida de Rol se lanza 200 veces un dado de cuatro caras obteniéndose 60 veces el número 1, 45 veces el número 2, 38 veces el número 3 y 57 veces el número 4. Se puede aceptar, a un nivel de confianza del 95%, que estos resultados corresponden a un dado homogéneo.

Solución:

1º La hipótesis nula será que el dado es homogéneo, esto implica que la distribución de los números es uniforme, es decir que los cuatro números tienen una probabilidad de aparecer de 0,25.

2º La hipótesis alternativa será que la distribución no es uniforme.

3º Como la variable es discreta utilizaremos el test Ji-cuadrado de bondad de ajuste a una distribución.

4º En la tabla siguiente se han realizado todos los cálculos necesarios, obteniéndose el valor 4,36 para el estadístico de contraste.

$x_i$	$n_i$	$p_i$	$Np_i$	$n_i - np_i$	$(n_i - np_i)^2$	$(n_i - np_i)^2 / np_i$
1	60	0,25	50	10	100	2
2	45	0,25	50	-5	25	0,5
3	38	0,25	50	-12	144	2,88
4	57	0,25	50	7	49	0,98
	200					4,36

5º Como el estadístico tenía 4 sumandos, buscamos en las tablas de la Ji-cuadrado con 3 grados de libertad el valor que deja por debajo una probabilidad de 0,95 y obtenemos que el valor crítico es 7,81.

6º Como el valor del estadístico es inferior al valor crítico, aceptamos la hipótesis nula.

7º Estos resultados son compatibles con el hecho de que el dado sea homogéneo.

2.- En una encuesta preelectoral realizada a 500 personas se obtuvo la siguiente distribución en función de sus edades y de su intención de voto:

Partido	Edad		
	18 – 35	35 – 50	50 o más
A	10	40	60
B	15	70	90
C	45	60	35
D	30	30	15

A un nivel de confianza del 90% ¿Puede afirmarse que la intención de voto es independiente de la edad?

Solución:

1º La hipótesis nula es que las dos variables son independientes.

2º La hipótesis alternativa es que hay relación entre ambas variables.

3º Se trata de un contraste de independencia entre dos variables, por consiguiente el estadístico de contraste a utilizar es el estadístico Ji-cuadrado para tablas de contingencia.

4º Las tablas siguientes presentan los cálculos del estadístico:

Partido	Edad			
	18 – 35	35 – 50	50 o más	
A	10	40	60	110
B	15	70	90	175
C	45	60	35	140
D	30	30	15	75
	100	200	200	500

A partir de las frecuencias marginales de la tabla anterior, se obtienen las frecuencias esperadas que aparecen a continuación:

Partido	Edad		
	18 – 35	35 – 50	50 o más
A	22	44	44
B	35	70	70
C	28	56	56
D	15	30	30

Por consiguiente las discrepancias entre frecuencias empíricas y frecuencias esperadas son:

Partido	Edad		
	18 – 35	35 – 50	50 o más
A	-12	-4	16
B	-20	0	20
C	17	4	-21
D	15	0	-15

Los cuadrados de las discrepancias son:

Partido	Edad		
	18 – 35	35 – 50	50 o más
A	144	16	256
B	400	0	400
C	289	16	441
D	225	0	225

Dividiendo por las frecuencias esperadas se obtiene:

Partido	Edad			
	18 – 35	35 – 50	50 o más	
A	6,55	0,36	5,82	
B	11,43	0	5,71	
C	10,32	0,29	7,88	
D	15	0	7,5	
	43,30	0,65	26,91	70,86

Sumando, se obtiene el valor del estadístico 70,86.

5º Como la edad presenta tres intervalos y los partidos son cuatro, el estadístico tendrá  $(3 - 1) \cdot (4 - 1) = 6$ . Buscamos en las tablas de la distribución Ji-cuadrado con 6 grados de libertad el valor de la variable que deja por debajo una probabilidad de 0,9 encontramos que el valor crítico es 10,64.

6º Como el valor del estadístico es mayor que el valor crítico rechazamos la hipótesis nula de que ambas variables son independientes.

7º La edad cambia la intención de voto.

3.- Los tiempos de respuesta de 9 sujetos en una tarea de reconocimiento de palabras, previamente presentadas, han sido los siguientes:

115, 98, 123, 109, 112, 87, 118, 104, 116

A un nivel de confianza del 95% ¿Son compatibles estos resultados con la hipótesis de que el tiempo de reacción en esta tarea sigue una distribución Normal de media 110 y desviación típica 10?

Solución:

1º La hipótesis nula es que los datos proceden de una Normal (110, 10).

2º La hipótesis alternativa es que no siguen esa distribución Normal.

3º Como la variable es continua, y la hipótesis nula especifica totalmente la distribución utilizaremos el test de Kolmogoroff-Smirnoff, cuyo estadístico de contraste es:

$$\max | F_n(x_i) - M_n(x_i) |$$

4º los cálculos del estadístico se especifican en la siguiente tabla:

$x_i$	87	98	104	109	112	115	116	118	123
$z_i$	-2,3	-1,2	-0,6	-0,1	0,2	0,5	0,6	0,8	1,3
$F_n$	0,0107	0,1151	0,2743	0,4602	0,5793	0,6915	0,7257	0,7881	0,9032
$M_n$	0,1111	0,2222	0,3333	0,4444	0,5556	0,6667	0,7778	0,8889	1
$ F_n - M_n $	0,1004	<b>0,1071</b>	0,059	0,0158	0,0237	0,0248	0,0521	0,1008	0,0968

5º Buscando en las tablas del test Kolmogoroff-Smirnoff para  $n = 9$  el valor crítico para un nivel de confianza del 95% se obtiene 0,43001.

6° Como el valor del estadístico 0,1071 es menor que el valor crítico se acepta la hipótesis nula.

7° A un nivel de confianza del 95% no hay evidencia en contra de que el tiempo de reacción siga una distribución  $N(110, 10)$ .

4.- En la encuesta telefónica realizada el pasado curso por los alumnos los resultados fueron muy dispares, mientras algunos realizaron las cuatro entrevistas programadas otros no consiguieron cumplimentar ninguna de ellas. La distribución del número de entrevistas conseguidas por los 57 alumnos que participaron en el proyecto fue la siguiente:

Nº entrevistas	Nº alumnos
0	6
1	16
2	24
3	9
4	2
Total	57

A un nivel de confianza del 90% ¿Puede afirmarse que estas diferencias han sido debidas al azar? O por el contrario están motivadas por alguna otra causa.

1° La hipótesis nula de que los resultados obtenidos son debidos al azar implica que en todas las llamadas hay la misma probabilidad de conseguir respuesta y que el resultado de cada llamada es independiente de las restantes. Entonces el número de entrevistas conseguidas por cada alumno es la suma de cuatro variables de Bernoulli y por consiguiente, la distribución sería una Binomial con  $n = 4$  y  $P$  desconocida.

2° La hipótesis alternativa es que no siguen esa distribución Binomial.

3° Como la variable es discreta y además la hipótesis nula no especifica totalmente la distribución utilizaremos el test Ji-cuadrado de bondad de ajuste.

4° Para calcular el valor del estadístico necesitamos las probabilidades de los valores, para ello es preciso estimar previamente el valor de  $P$  a partir de los datos de la muestra.

El total de llamadas ha sido  $57 \cdot 4 = 228$ .

Las llamadas con éxito han sido  $1 \cdot 16 + 2 \cdot 24 + 3 \cdot 9 + 4 \cdot 2 = 99$ .

La proporción es  $99/228 = 0,4342$ .

Las probabilidades aplicando la función de probabilidad de la distribución Binomial serán:

$$\Pr(x = 0) = \binom{4}{0} 0,4342^0 \cdot 0,5658^4 = 0,1025$$

$$\Pr(x = 1) = \binom{4}{1} 0,4342^1 \cdot 0,5658^3 = 0,3146$$

$$\Pr(x = 2) = \binom{4}{2} 0,4342^2 \cdot 0,5658^2 = 0,3621$$

$$\Pr(x = 3) = \binom{4}{3} 0,4342^3 \cdot 0,5658^1 = 0,1853$$

$$\Pr(x = 4) = \binom{4}{4} 0,4342^4 \cdot 0,5658^0 = 0,0355$$

El resto de los cálculos necesarios para obtener el valor del estadístico aparecen ordenados en la tabla siguiente:

$x_i$	$n_i$	$p_i$	$np_i$	$n_i - np_i$	$(n_i - np_i)^2$	$(n_i - np_i)^2 / np_i$
0	6	0,1025	5,84	0,16	0,03	0,005
1	16	0,3146	17,93	-1,93	3,72	0,207
2	24	0,3621	20,64	3,36	11,29	0,547
3	9	0,1853	10,56	-1,56	2,43	0,230
4	2	0,0355	2,02	-0,02	0	0
	57					0,989

5° El estadístico tiene 5 sumandos, pero como hemos estimado un parámetro, debemos buscar en las tablas de la Ji-cuadrado con 3 grados de libertad. El valor de la variable que deja por debajo una probabilidad de 0,9 es 6,25.

6° Como el valor del estadístico 0,989 es menor que el valor crítico, 6,25 se acepta la hipótesis nula.

7° Los resultados obtenidos por los alumnos pueden ser fruto del azar.