

Estadística II
Grado en Matemáticas, UAM, 2019-2020

Hoja 2 (Contrastes no paramétricos, parte 1)

CONTRASTE χ^2 DE MODELO

1. Da una demostración directa para el caso $k = 2$ de que la distribución del estadístico del contraste χ^2 de bondad de ajuste converge a una distribución χ_1^2 es decir,

$$T = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} \rightarrow_d \chi_1^2 \quad \text{si } n \rightarrow \infty.$$

(Indicación: Hay que demostrar que $T = X_n^2$, donde $X_n \sim N(0, 1)$. Para reducir los dos sumandos a uno, utilizar la relación existente entre O_1 y E_2 y O_2, E_2 .)

2. Un cierto programa de ordenador se supone que genera cifras de cero a nueve totalmente al azar. En las 200 cifras obtenidas se observaron las siguientes frecuencias:

Números	0	1	2	3	4	5	6	7	8	9
Frecuencia	27	22	15	18	23	26	19	20	16	14

Con un nivel de significación de 0.05, ¿es “bueno” el programa?

3. Para comprobar el funcionamiento de un generador de muestras de la $N(0, 1)$, obtenemos la siguiente muestra de tamaño 450:

Frecuencia	Rango
30	menores que -2 ;
80	entre -2 y -1 ;
140	entre -1 y 0 ;
110	entre 0 y 1 ;
60	entre 1 y 2 ;
30	mayores que 2 .

¿Se puede aceptar, al nivel $\alpha = 1\%$, que el programa funciona correctamente?

4. Para estudiar el número de ejemplares de cierta especie en peligro de extinción que viven en un bosque, se divide el mapa del bosque en nueve zonas y se cuenta el número de ejemplares de cada zona. Se observa que 60 ejemplares viven en el bosque repartidos en las 9 zonas de la siguiente forma:

8	7	3
5	9	11
6	4	7

Mediante un contraste de hipótesis, analiza si estos datos aportan evidencia empírica de que los animales tienen tendencia a ocupar unas zonas del bosque más que otras.

5. Se desea estudiar el número de accidentes por día que se producen en cierto regimiento. Para ello se toman al azar los partes de 200 días dentro de los últimos 5 años, encontrando los siguientes resultados:

Número de accidentes en el día	0	1	2	3	4	5	6
Número de días	58	75	44	18	3	1	1

(a) ¿Se puede aceptar, con nivel de confianza del 90 %, que el número de accidentes por día sigue una distribución de Poisson?

(b) Asumiendo que el número de accidentes por día sigue una $\text{Poisson}(\lambda)$, ¿hay suficiente evidencia estadística (con nivel de significación $\alpha = 5\%$) de que el verdadero valor medio λ del número de accidentes por día es menor que 1.35? ¿El p -valor es mayor o es menor que 5 %?

6. Se clasificaron 1000 individuos de una población según el sexo y según tuvieran visión normal o fueran daltónicos.

	Masculino	Femenino
Normal	442	514
Daltónicos	38	6

Según un modelo genético, las probabilidades deberían ser:

$$\begin{array}{|c|c|} \hline \frac{1}{2}p & \frac{1}{2}p^2 + pq \\ \hline \frac{1}{2}q & \frac{1}{2}q^2 \\ \hline \end{array}$$

donde $q = 1 - p$ = proporción de genes defectuosos en la población.

(a) Estima el parámetro p a partir de la muestra (solución: $\hat{p} = 91.29\%$).

(b) ¿Concuerdan los datos con el modelo?

7. El número de asesinatos cometidos en Nueva Jersey cada día de la semana durante el año 2003 se muestra en la tabla siguiente:

Día	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Frecuencia	42	51	45	36	37	65	53

(a) Contrasta al nivel $\alpha = 0.05$, mediante un test χ^2 , la hipótesis nula de que la probabilidad de que se cometa un asesinato es la misma todos los días de la semana.

(b) Contrasta la hipótesis nula de que la probabilidad de que se cometa un asesinato es la misma desde el lunes hasta el viernes, y también es la misma los dos días del fin de semana (pero no es necesariamente igual en fin de semana que de lunes a viernes).

8. Según los datos de un exhaustivo estudio de mercado que se realizó en una gran ciudad, las ventas de impresoras para ordenadores personales de uso doméstico se dividen entre cuatro marcas (A, B, C y D) cuyos porcentajes del total de las ventas son 18 %, 22 %, 35 % y 25 %, respectivamente.

Un año después, se quiere analizar de nuevo la situación pero no se dispone de dinero para repetir un estudio de mercado a gran escala. Se decide, por tanto, observar la marca adquirida por 200 compradores de impresoras elegidos al azar, obteniéndose que de ellos 28 habían elegido la marca A, 48 la B, 77 la C y 47 la D.

¿Hay suficiente evidencia estadística, al nivel 5 %, para afirmar que el reparto del mercado ya no es el mismo que el año anterior?

OTROS CONTRASTES χ^2

9. Se quiere estudiar la relación entre la edad de los chicos y el tiempo T que ven semanalmente la televisión. Con una muestra de 200 chicos se obtuvieron los siguientes resultados:

CHICOS	menos de 14 horas	entre 14 y 18 horas	más de 18 horas
entre 6 y 10 años	20	30	30
entre 11 y 15 años	20	40	60

¿Hay evidencia estadística significativa con nivel $\alpha = 0.05$ de que existe relación entre la edad y el tiempo que ven la televisión?

10. En un estudio sobre el uso de sistemas operativos se han seleccionado al azar 150 profesores universitarios (PU), 150 profesionales técnicos (PT) de grado medio que trabajan en la industria y 150 personas con cargos directivos (CD) en empresas. A cada uno de los seleccionados se le ha preguntado cuál es el sistema operativo (A, B o C) que utiliza habitualmente en su trabajo con ordenadores. Los resultados se resumen en la siguiente tabla:

	A	B	C
PU	52	40	58
PT	42	45	63
CD	28	47	75

¿Hay suficiente evidencia estadística, al nivel 0.05, para concluir que existe alguna asociación entre el status profesional y la preferencia por un sistema operativo?

11. Se quiere comparar la biodiversidad de dos montes cercanos. En uno de los montes se eligen al azar 50 zonas, de 4 m² cada una, y se hace el recuento del número de especies vegetales diferentes que hay en cada una, con los resultados de la tabla de la izquierda. En el otro monte se hace el mismo recuento en otras 40 zonas, obteniéndose los resultados de la tabla de la derecha:

Número de zonas	Número de especies	Número de zonas	Número de especies
20	menos de 6	12	menos de 6
17	entre 6 y 8	20	entre 6 y 8
13	más de 8	8	más de 8

¿Son similares los dos montes en lo que se refiere a su biodiversidad? Haz el contraste correspondiente con un nivel de significación del 10 %.

12. Una fábrica de automóviles quiere averiguar si la preferencia de modelo tiene relación con el sexo de los clientes. Se toman dos muestras aleatorias de 1000 hombres y 1000 mujeres y se observan las siguientes preferencias:

SEXO	MODELO		
	A	B	C
Mujer	340	400	260
Hombre	350	270	380

¿Son homogéneas las preferencias entre hombres y mujeres, al nivel de significación 0.01?

Supongamos ahora que los resultados de la tabla anterior correspondiesen a una muestra aleatoria de 2000 personas, clasificadas de acuerdo con el doble criterio sexo/preferencia de modelo. ¿En qué se modificaría el planteamiento o la interpretación de los resultados anteriores?

13. Un estudio sobre tabaquismo en tres comunidades, mediante tres muestras aleatorias de tamaño 100, proporciona los siguientes resultados:

Comunidad	fumadores	no fumadores
A	13	87
B	17	83
C	18	82

¿Pueden considerarse homogéneas las tres poblaciones en cuanto a sus hábitos fumadores?

14. Se ha clasificado una muestra aleatoria de 500 hogares de acuerdo con su situación en la ciudad (Sur o Norte) y su nivel de renta (en miles de euros) con los siguientes resultados:

Renta	Sur	Norte
0 a 10	42	53
10 a 20	55	90
20 a 30	47	88
más de 30	36	89

(a) A partir de los datos anteriores, contrasta a nivel $\alpha = 0.05$ la hipótesis nula de que en el sur los hogares se distribuyen uniformemente en los cuatro intervalos de renta considerados.

(b) A partir de los datos anteriores, ¿podemos afirmar a nivel $\alpha = 0.05$ que la renta de los hogares es independiente de su situación en la ciudad?

15. Se ha realizado una encuesta para estudiar posibles relaciones entre el nivel educativo (educación superior, media o primaria) de las personas y el nivel de consumo (bajo, medio o alto) de un determinado producto. Los resultados, para 400 personas seleccionadas al azar, han sido:

	BAJO	MEDIO	ALTO
SUPERIOR	31	41	44
MEDIA	28	79	125
PRIMARIA	16	17	19

Contrasta estadísticamente (nivel 0.01) la independencia entre nivel educativo y nivel de consumo.

16. Se desea evaluar la efectividad de una nueva vacuna antigripal. Para ello se decide suministrar dicha vacuna, de manera voluntaria y gratuita, a una pequeña comunidad. La vacuna se administra en dos dosis, separadas por un período de dos semanas, de forma que algunas personas han recibido una sola dosis, otras han recibido las dos y otras personas no han recibido ninguna. La siguiente tabla indica los resultados que se registraron durante la siguiente primavera en 1000 habitantes de la comunidad elegidos al azar.

	No vacunados	Una dosis	Dos dosis
Gripe	24	9	13
No gripe	289	100	565

¿Proporcionan estos datos suficiente evidencia estadística (al nivel de significación 5%) para indicar una dependencia entre la clasificación respecto a la vacuna y la protección frente a la gripe?

17. Se ha llevado a cabo una encuesta a 100 hombres y 100 mujeres sobre su intención de voto. De las 100 mujeres, 34 quieren votar al partido A y 66 al partido B. De los 100 hombres, 50 quieren votar al partido A y 50 al partido B.

(a) Utiliza un contraste basado en la distribución χ^2 para determinar si con estos datos se puede afirmar a nivel $\alpha = 0.05$ que el sexo es independiente de la intención de voto.

(b) Determina el intervalo de valores de α para los que la hipótesis de independencia se puede rechazar con el contraste del apartado anterior.