Matemàtiques II. Full 12 d'exercicis per entregar. Solucions.

1) En un estudio del 2003 se investigó si el historial familiar del paciente con desorden bipolar tiene influencia en la edad en que se le manifiesta este desorden. Se tomó un grupo de enfermos al azar y se anotó su historial familiar y la edad en que se manifestó la dolencia, clasificándolos en este punto en Precoces (en los cuales el desorden bipolar se manifestó antes del 18 años) y Tardíos (en los cuales el desorden bipolar se manifestó después del 18 años). Los resultados son los de la tabla siguiente:

Historial Familiar	Precoces	Tardíos
Negativo	28	35
Desorden bipolar	19	38
Desorden unipolar	41	44
Desórdenes unipolar y bipolar	53	60

Como decimos, queremos contrastar si el historial familiar influye en la edad en la cual se manifestó el desorden bipolar en el paciente.

a) Realizáis a mano y en condiciones de control (sin emplear R) una maceta χ^2 sobre estos datos para contrastar si las frecuencias de los diferentes tipos de desorden difieren según la edad en la cual se manifestó el desorden bipolar en el paciente. Donau los detalles de los cálculos.

Empezamos calculando las marginales:

Historial Familiar	Precoces	Tardíos	Suma
Negativo	28	35	63
Desorden bipolar	19	38	57
Desorden unipolar	41	44	85
Desórdenes unipolar y bipolar	53	60	113
Suma	141	177	318

Calculamos las frecuencias esperadas:

Historial Familiar	Precoces	Tardíos
Negativo	27.93	35.07
Desorden bipolar	25.27	31.73
Desorden unipolar	37.69	47.31
Desórdenes unipolar y bipolar	50.1	62.9

Calculamos la tabla de entradas (freq. observada-freq. esperada)^2/freq. esperada:

Historial Familiar	Precoces	Tardíos
Negativo	0.0002	0.0001
Desorden bipolar	1.5557	1.239
Desorden unipolar	0.2907	0.2316
Desórdenes unipolar y bipolar	0.1679	0.1337

El estadístico X^2 será la suma de las entradas de esta mesa

$$X^2 = 0.0002 + 0.0001 + 1.5557 + 1.239 + 0.2907 + 0.2316 + 0.1679 + 0.1337 = 3.6189$$

Este estadístico sigue una ley χ^2 con $(2-1)(4-1) = 3$ grados de libertad, por lo tanto el p-valor es

$$P(\chi_3^2 \geq 3.6189) = 1 - P(\chi_3^2 \leq 3.6189) \approx 1 - 0.7 = 0.3$$

y por tanto no podemos rechazar la hipótesis nul·la que la edad en que se manifiesta la dolencia sea independiente del historial familiar.

b) Repetís esta maceta con la función pertinente de R. Procuráis que dé el mismo.

```
> frbip=matrix(c(28,35,19,38,41,44,53,60),nrow=4,byrow=TRUE)
> chisq.maceta(frbip)
      Pearson's Chi-squared maceta
data:  frbip
X-squared = 3.6216, df = 3, p-value = 0.3053
```

Los resultados son similares, y la conclusión la misma.

c) Según el diseño del experimento, qué tipo de maceta habéis realizado: de independencia o de homogeneidad? Justificáis vuestra respuesta.

A pesar de que desde el punto de vista del contraste son el mismo, este es una maceta de independencia: tomamos una muestra aleatoria de la población (enfermos con desorden bipolar), mesuramos las dos variables sobre cada individuo, y comprobamos si estas dos variables son independientes. En un macetas de homogeneidad, hubiéramos fijado una variable y hubiéramos tomado una muestra aleatoria de cada nivel de esta variable, y el conjunto de individuos obtenido de este modo no tendria por qué ser una muestra aleatoria del total de la población.

2) En un estudio se quiso contrastar si había relación entre el grupo sanguíneo de una persona y el hecho que sea o no portador de un cierto antígeno raro. Para hacerlo, se eligieron 150 portadores del antígeno y 500 no portadores y se los miró el grupo sanguíneo. Los resultados son los de la tabla siguiente:

Grupo	Portadores	No portadores
0	72	230
A 54		192
B	16	63
AB	8	15

a) Realizáis a mano y en condiciones de control (sin emplear R) una maceta χ^2 sobre estos datos para contrastar si las frecuencias de los diferentes tipos sanguíneos son diferentes en los portadores y los no portadores. Donau los detalles de los cálculos.

Empezamos calculando las marginales:

Grupo	Portadores	No portadores	Suma
0	72	230	302
A 54		192	246
B	16	63	79
AB	8	15	23
Suma	150	500	650

Calculamos las frecuencias esperadas:

Grupo	Portadores	No portadores
0	69.69	232.31
A 56.77		189.23
B	18.23	60.77
AB	5.31	17.69

Calculamos la tabla de entradas (freq. observada-freq. esperada)^2/freq. esperada:

Grupo	Portadores	No portadores
0	0.0766	0.023
A 0.1352		0.0405
B	0.2728	0.0818
AB	1.3627	0.4091

El estadístico X^2 será la suma de las entradas de esta mesa

$$X^2 = 0.0766 + 0.023 + 0.1352 + 0.0405 + 0.2728 + 0.0818 + 1.3627 + 0.4091 = 2.4017$$

Este estadístico sigue una ley χ^2 con $(2-1)(4-1) = 3$ grados de libertad, por lo tanto el p-valor es

$$P(\chi_3^2 \geq 2.4017) = 1 - P(\chi_3^2 \leq 2.4017) \approx 1 - 0.5 = 0.5$$

y por tanto no podemos rechazar la hipótesis nul·la que el grupo sanguíneo y ser o no portador del antígeno sean independientes

b) Repetís esta maceta con la función pertinente de R. Procuráis que dé el mismo.

```
> freqsang=matrix(c(72,230,54,192,16,63,8,15),nrow=4,byrow=TRUE)
> chisq.maceta(freqsang)
      Pearson's Chi-squared maceta
data:  freqsang
X-squared = 2.4052, df = 3, p-value = 0.4927
```

Los resultados son similares, y la conclusión la misma.

c) Según el diseño del experimento, qué tipo de maceta habéis realizado: de independencia o de homogeneidad? Justificáis vuestra respuesta.

A pesar de que desde el punto de vista del contraste son el mismo, este es una maceta de homogeneidad: fijamos la variable "ser portador", tomamos una muestra aleatoria de individuos de cada uno de sus dos niveles y mesuramos la otra variable sobre estos individuos. Fijaos que el conjunto de individuos elegido de este modo no tiene por qué ser una muestra aleatoria del total de la población.

3) La tabla siguiente da la distribución de la población de los EE. UU. por edades

Edades	Menos de 18	18–24	25–34	35–44	45–64	65–79	80 o más
Porcentaje %	25	9.6	13.4	14.4	25.4	9.4	2.8

Hemos estudiado una muestra de 130 habitantes de los EE. UU. sin seguro médico, y los resultados han sido los siguientes:

Edades	Menos de 18	18–24	25–34	35–44	45–64	65–79	80 o más
Frecuencia	30	14	18	22	40	5	1

a) Realizáis a mano y en condiciones de control (sin emplear R) una maceta para contrastar si hay evidencia con nivel de significación 0.01 que la distribución por edades de los habitantes de los EE. UU. sin seguro médico es diferente de la distribución por edades de todos los habitantes de los EE. UU..

Haremos una maceta χ^2. Empezamos calculando las frecuencias esperadas

Edades	Menos de 18	18–24	25–34	35–44	45–64	65–79	80 o más
Freq. Obs.	30	14	18	22	40	5	1
Prob. Esp.	0.25	0.096	0.134	0.144	0.254	0.094	0.028
Freq. Esp.	32.5	12.48	17.42	18.72	33.02	12.22	3.64

Como que la frecuencia esperada de la última clase es inferior a 5, lo agrupamos con el anterior.

Edades	Menos de 18	18–24	25–34	35–44	45–64	65 o más
Freq. Obs.	30	14	18	22	40	6
Freq. Esp.	32.5	12.48	17.42	18.72	33.02	15.86
(Freq.Obs. − Freq.Esp.)^2/Freq.Esp.	0.192	0.185	0.019	0.575	1.475	6.130

Calculamos el estadístico de contraste

$$X^2 = 0.192 + 0.185 + 0.019 + 0.575 + 1.475 + 6.13 = 8.576$$

La distribución de este estadístico es χ^2_5 y por tanto el p-valor es

$$P(\chi^2_5 \geq 8.576) = 1 - P(\chi^2_5 \leq 8.576) \approx 1 - 0.85 = 0.15$$

Esto muestra que no hay evidencia con nivel de significación 0.01 que las dos distribuciones sean diferentes.

b) Repetís esta maceta con la función pertinente de R. Procuráis que dé el mismo.

```
> freqobs=c(30,14,18,22,40,5,1)
> probs=c(0.25,0.096,0.134,0.144,0.254,0.094,0.028)
> freq.esp=probs*130
> freq.esp
[1] 32.50 12.48 17.42 18.72 33.02 12.22  3.64
> \#Tenemos que agrupar los dos últimos
> freqobs=c(30,14,18,22,5+1)
> probs=c(0.25,0.096,0.134,0.144,0.254,0.094+0.028)
> chisq.maceta(freqobs,p=probs)
     Chi-squared maceta for given probabilities
fecha:  freqobs
X-squared = 8.5768, df = 5, p-value = 0.1272
```

Los resultados son similares, y la conclusión la misma.