Ejercicios T.6 CONTRASTES NO PARAMÉTRICOS

1. Se quiere comprobar si el número de asignaturas aprobadas en una determinada convocatoria universitaria sigue una distribución de Poisson (nivel de significación 0,05; para ello disponemos de la siguiente información de 60 alumnos:

número de aprobadas	nº alumnos
0	10
1	15
2	15
3	10
4	6
5	4

Estimamos que la distribución de poisson tendrá de media quedando la hipótesis H_0 : el número de aprobadas sigue una Poisson de media 1,9833
(ir a script de poisson)
realizamos un contraste de la chi-dos, así: (ir a script de contraste de chi-2)

aprobadas	$n_{0,i}$	$P(x_i)=\frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$	$n_{T,i}=P(x_i)\cdot N$	$\left(n_{o,i}-n_{t,i}\right)^2$	$\frac{\left(n_{o,i}-n_{t,i}\right)^2}{n_{t,i}}$
0	10	0,1376	8,26	3,027	0,3664
1	15	0,272	16,32	1,742	0,1067
2	15	0,270	16,2	1,44	0,0888
3	10	0,1789	10,73	0,5329	0,0496
4	6 (10)	0,088	5,3 (7,4)	6,76	0.9135
5	4	0,0351	2,106		
					χ^2=1,525

Dado que hay una frecuencia teórica menor que 5 agrupamos para 4 y 5 aprobadas El estadístico seguirá una χ^2 con m-k-1=5-1-1=3 gl. Cuyo valor según tabla (ir a tabla de la χ^2) para un nivel de significación del 5% será = 7,815. Dado que 1,525<7,815 No rechazamos la hipótesis de que las asignaturas aprobadas e se distribuyan según una Poisson.

2. Realizar este mismo ejercicio utilizando el test de Kolmogorov-Smirnov.

Así:
Tendremos que construir la tabla de frecuencias relativas acumuladas, tanto teóricas como observadas: así

apr	$n_{0,i}$	$P(x_i) = \frac{e^{-\lambda} \cdot \lambda^{x_i}}{x_i!}$	$F_o(x_i)$	$F_T(x_i)$	$D = \lvert F_t(x_i) - F_o(x_i) \rvert$
0	10	0,1376	10/60=0,16	0,13	0,03 máximo
1	15	0,272	25/60=0,41	0,4	0,01
2	15	0,270	40/60=0,66	0,67	0,01
3	10	0,1789	50/60=0,83	0,85	0,02
4	6	0,088	56/60=0,93	0,94	0.01
5	4	0,0351	1	0,98	0,02

el máximo resultado de la diferencia es 0,03 que será el estadístico a comparar en la *tabla de K-S para un nivel de significación de 0.05 y n=60 tenemos que:*

—

(ir a tabla de K-S) dado que es mayor que 0,03 no rechazamos la hipótesis de que los aprobados siguen una distribución de Poisson. (ir a tets de K-S)

---

3. En nuestra empresa tenemos organizados tres turnos de producción. Hemos recogido información aleatoria sobre las cantidades de piezas que se producen por turno a lo largo de 100 días, resultando la siguiente información recogida en la tabla.

Turno/producción	mañana	tarde	noche
100-200 piezas	12	12	13
200-300 piezas	7	12	12
300-400 piezas	10	8	14

Contrastar con un nivel de significación del 2% si la producción de piezas en los diversos turnos es homogénea.

Se trata de un contraste de homogeneidad luego (ir a script de contraste de homogeneidad)

H_0 :Comportamiento homogéneo de las subpoblaciones "turno" al respecto de la variable piezas producidas

Estableceremos la siguiente tabla donde aparecen las frecuencias conjuntas observadas en la muestra y las que teóricamente corresponderían al caso de homogeneidad. Así

Turno/producción	mañana		tardes		noche		
	obs	teo	obs	teo	obs	teo	
100-200 piezas	12	10,73	12	11,84	13	14,43	37
		0,15		**0,0021**		**0,14**	
200-300 piezas	7	8,99	12	9,92	12	12,09	31
		0,44		**0,43**		**0,00**	
300-400 piezas	10	9,28	8	10,24	14	12,48	32
	0,05		**0,49**		**0,18**		
	29		32		39		100

$$n_{i,j}^{T} = \frac{n_{\cdot j} \cdot n_{i\cdot}}{N} \rightarrow n_{1,1}^{T} = \frac{n_{\cdot 1} \cdot n_{1,\cdot}}{N} = \frac{29 \cdot 37}{100} = 10{,}73$$

y de la misma manera para todas las frecuencias conjuntas teóricas

cada uno de estas celdas sería

$$\frac{\left(n_{o,i} - n_{t,i}\right)^2}{n_{t,i}}$$

en este caso

$$\frac{\left(7 - 8{,}99\right)^2}{8{,}99} = 0{,}44$$

y así para todas las diferencias entre observada y teóricas y siendo el estadístico:

$$\chi^2 = \sum_{\forall i} \sum_{\forall j} \frac{\left(n_{o,i} - n_{t,i}\right)^2}{n_{t,i}}$$

Cuyo valor sería =0,15+0,0021+0,14+0,44+0,43+0+0,05+0,49+0,18=**1,88**

el estadístico se distribuye según una χ^2 con (m-1)·(n-1) gl. es decir con (3-1)(3-1)=4gl para un nivel de significación de 0,02 (ir a tabla de laχ^2) tomará el valor 11,668: dado que el valor del estadístico 1,88 es menor que 11,668 No rechazamos la hipótesis de comportamiento homogéneo de los turnos.

---

4. Deseamos conocer si existe relación entre la edad de nuestros empleados y su rendimiento profesional; para ello hemos realizado pruebas a 100 de ellos con los resultados siguientes:

rendimiento\ edad	menores de 45	mayores de 45
óptimo	12	30
medio	18	40

contrastar dicha hipótesis con nivel de confianza del 94%
(ir a script de contraste independencia)

El problema radica en realizar un contraste de independencia entre la edad y el rendimiento, si esto es así (independencia) no existirá relación entre ambas, luego:

H_0 : existe independencia entre edad y rendimiento
para realizarlo construimos la tabla de contingencia que incluya frecuencias observadas y teóricas:

rendimiento\edad	menores 45		mayores 45		
	obs	teo	obs	teo	
óptimo	12	12,6	30	29,4	42
		0,0008		**0,0003**	
medio	18	17,4	40	40,6	58
		0,0006		**0,0002**	
	30		70		100

dado que la distribución del estadístico será (Ir a otro script de solución)

χ^2 con (m-1)·(n-1) gl. es decir con (2-1)(2-1)=1gl es de aplicación la corrección de yates:

calculemos primero las frecuencias conjuntas teóricas.

$$n^{T}_{i,j} = \frac{n_{\cdot j}\cdot n_{i\cdot}}{N} \rightarrow n^{T}_{1,1} = \frac{n_{\cdot 1}\cdot n_{1,\cdot}}{N} = \frac{30\cdot 42}{100} = 12{,}6$$

y de la misma manera para todas las frecuencias conjuntas teóricas

dado que es de aplicación la corrección de Yates tendremos que cada celda será:

$$\frac{\left(\left|n^{T}_{ij} - n^{o}_{ij}\right| - 0{,}5\right)^{2}}{n^{T}_{ij}} \rightarrow$$

en ese caso

$$\frac{\left(\left|12{,}6 - 12\right| - 0{,}5\right)^{2}}{12{,}6} \cong 0{,}0008$$

y así para cada par de frecuencias; teórica y observada siendo el estadístico:

$$\chi^{2} = \sum_{\forall i}\sum_{\forall j} \frac{\left(\left|n^{T}_{ij} - n^{o}_{ij}\right| - 0{,}5\right)^{2}}{n^{T}_{ij}} = 0{,}0008 + 0{,}0003 + 0{,}0002 + 0{,}0006 = 0{,}0019$$

el estadístico se distribuye como dijimos según una χ^{2} con 1gl que para α =0,06 toma el valor según tabla (ir a tabla de laχ^{2}) de 3,537. Dado que este valor es mayor que 0,0019. No podemos rechazar la hipótesis de independencia luego la edad no influye en el rendimiento laboral.

---

5. Se realizó una encuesta a 80 hogares visitados antes y después de una campaña publicitaria, construyéndose la siguiente tabla de viviendas que tenían el producto anunciado.

producto\campaña	antes	después
tienen	6	16
no tienen	25	33

Contrastar con un nivel de significación del 1% si la campaña ha sido eficaz

La campaña es eficaz si el comportamiento del producto es distinto (mejor) después que antes de la campaña, estamos, pues, ante un contraste de homogeneidad: (Ir a script de contraste de homogeneidad)

H_0 :Homogeneidad de comportamiento del producto antes y después de la campaña: así:

	antes		después		
	obs	**teo**	**obs**	**teo**	
tienen	**6**	**8,52**	**16**	**13,5**	**22**
		0,47		**0,296**	
no tienen	**25**	**22,5**	**33**	**35,5**	**58**
		0,177		**0,112**	
$n_{*,j}$	**31**		**49**		**80**

dado que la distribución del estadístico será   ir a script de solución

χ^2 con (m-1)·(n-1) gl. es decir con (2-1)(2-1)=1gl es de aplicación la corrección de Yates:

calculemos primero las frecuencias conjuntas teóricas.

$$n_{i,j}^{T} = \frac{n_{\cdot j} \cdot n_{i\cdot}}{N} \rightarrow n_{1,1}^{T} = \frac{n_{\cdot 1} \cdot n_{1\cdot}}{N} = \frac{22 \cdot 31}{80} = 8{,}52$$

y de la misma manera para todas las frecuencias conjuntas teóricas

dado que es de aplicación la corrección de Yates tendremos que cada celda será:

$$\frac{\left(\left|n_{ij}^{T} - n_{ij}^{O}\right| - 0{,}5\right)^2}{n_{ij}^{T}} \rightarrow$$

en ese caso

$$\frac{\left(\left|8{,}52 - 6\right| - 0{,}5\right)^2}{8{,}52} \cong 0{,}47$$

y así para cada par de frecuencias; teórica y observada siendo el estadístico:

$$\chi^2 = \sum_{\forall i} \sum_{\forall j} \frac{\left(\left|n_{ij}^{T} - n_{ij}^{O}\right| - 0{,}5\right)^2}{n_{ij}^{T}} = 0{,}47 + 0{,}296 + 0{,}177 + 0{,}112 = 1{,}055$$

el estadístico se distribuye como dijimos según una χ^2 con 1gl que para α =0,01 toma el valor según tabla de 6,635. (ir a tabla de la χ^2) Dado que este valor es mayor que 1,055 No podemos rechazar la hipótesis de homogeneidad luego la campaña no ha sido eficaz .

6. Una muestra de 200 empleados no cualificados de una determinada actividad manufacturera nos ha permitido construir la siguiente tabla de frecuencias conjuntas de antigüedad y salarios. Contrastar con un nivel de significación del 1% la posible independencia entre ambas características.

antigüedad/salarios	poca	media	mucha
bajos	62	10	2
medios	14	38	12
altos	2	9	51

contraste de independencia con la Jhi-2 con (n-1)(m-1) g.l = 2x2= 4 no es de aplicación Yates.

El estadístico será

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\left(n_{ijT}-n_{ijO}\right)^2}{n_{ijT}}$$

donde las observadas vienen dadas por la tabla y las teóricas resultan de aplicar

$$n_{ijT}=\frac{n_{i,.} * n_{.,j}}{N}$$

así de manera resumida en la tabla (ir a script de contraste de independencia)

antigüedad/salarios	poca		media		mucha		$n_{i,.}$
	Ob	Teo	Ob	Teo	Ob	Teo	
bajos	62	28,86	10	24,05	2	24,05	74
medios	14	24,96	38	18,24	12	20,8	64
altos	2	24,28	9	17,67	51	20,15	62
$n._j$	78		57		65		200

el estadístico χ^2 = 38,05+5,83+20,21+4,81+21,40+3,72+20,34+4,25+47,23=165,84

en la tabla (ir a tabla de la χ^2) $\chi^2_{4,0,01}$=13,28 el estadístico es mayor: luego rechazamos la hipótesis de independencia.