

16

Bondad de ajuste

En el capítulo anterior se mostró la primera parte del análisis de frecuencias y se mencionó el empleo de la prueba de Chi-cuadrado y la prueba de G, para determinar si los valores observados se ajustaban a los esperados. En este tipo de ensayos se busca establecer si las diferencias observadas entre los datos reales y los teóricos se deben al azar, o si por el contrario la teoría no es buena para explicar la realidad. A eso se lo denomina ensayos de *bondad de ajuste*. Los problemas vistos cuando las frecuencias esperadas se calculaban con la teoría mendeliana, son en esencia casos de bondad de ajuste. En cambio, en los problemas presentados como tablas de contingencia, se buscaba probar la *independencia* de los factores involucrados entre sí. En este capítulo se tratarán con más detalle los casos de ajuste. El caso clásico y más difundido es la prueba de Chi-cuadrado. Desde 1900 y a través de su revista *Biometrika*, Pearson difundió su prueba ampliamente. El ejemplo más famoso fue el informe de Student sobre el error de recuento en hemocitómetro, donde proponía ajustar con una Poisson los valores observados, en lugar de usar la acostumbrada curva de Gauss. En las últimas décadas la prueba de la G se está imponiendo por su mayor sensibilidad y robustez. Se constituye en la opción más recomendable para el caso paramétrico. Cuando los supuestos no se cumplen, o sencillamente por su facilidad de ejecución, la prueba más recomendable en el campo no paramétrico es la de Kolgomorov-Smirnov, para una y para comparar dos muestras entre sí. Todos estos modelos se presentan a continuación.

16.1 El método clásico de Pearson

Este método aplica el test de la distribución de Pearson: la Chi cuadrado. Consiste en comparar las frecuencias observadas en un experimento, con sus respectivas frecuencias observadas, obtenidas con la distribución teórica que explica el fenómeno estudiado. Hasta la primera década del siglo XX, la distribución Gaussiana era el enfoque aceptado por todos para explicar la Teoría de errores en mediciones. El punto de inflexión histórico a esa tendencia se produce en 1907, con el trabajo de un ingeniero químico W. Gosset, quien en su tarea diaria efectuaba conteos de células de levadura de cerveza en una cámara de recuento de Neubauer denominada hemocitómetro. Gosset sospechó que sus datos se ajustaban mejor a una distribución Poisson que a la de Gauss y para probarlo usó el método de Pearson. Publicó sus trabajos en la revista *Biometrika* dirigida por el mismo Pearson, su mentor en cuestiones matemáticas. Estas publicaciones llegaron a manos de un estudiante avanzado escocés: R.A. Fisher, quien usó un comentario de Gosset respecto a la posible existencia de una distribución teórica, mejor que la de Gauss en

muestras pequeñas, para desarrollar un nuevo modelo: el de Student. Allí comenzó la moderna estadística, con aplicaciones innovadoras en el trabajo experimental. Parece adecuado comenzar este tema con ese caso. En la Tabla 16.1 se presentan los datos de Student (1907) para mostrar el uso del ajuste a una función de Poisson:

Tabla 16.1 El método de la Chi cuadrado.

N	O _i	E _i	O* _i	E* _i	$\chi^2=(O^*i-E^*i)^2/E^*i$	χ^2_{corr}
0	75	66,12	75	66,12	1,193	1,062
1	103	119,02	103	119,02	2,156	2,024
2	121	107,11	121	107,11	1,801	1,674
3	54	64,27	54	64,27	1,641	1,485
4	30	28,92	30	28,92	0,040	0,012
5	13	10,41	17	14,56	0,409	0,258
6	2	3,12				
7	1	0,80				
8	0	0,18				
9 y más	1	0,05				
Total	400	400	400	400	7,240	6,515

N : número de células dentro de cada cuadrícula.
O_i frecuencia observada (cantidad de cuadrículas con N células adentro).
E_i : frecuencia esperada obtenida con la fórmula de Poisson.
E*_i : frecuencia agrupada.

Si la frecuencia observada es muy chica ($O_i < 5$), los valores de frecuencia esperada se distorsionan mucho y los valores de χ^2 respectivos contribuyen demasiado al total, pudiendo mostrar significación donde en realidad no la hay. Una regla práctica establecida hace muchos años es agrupar las frecuencias anexas (cuando $O_i < 5$) en un valor manejable. En la Tabla 16.1 como las cuatro últimas frecuencias observadas son chicas, se agrupan con la más cercana, lo mismo que con las esperadas. Entonces, los datos agrupados se muestran en las columnas cuarta y quinta de la tabla ($O^*6 = 13+2+1+0+1 = 17$ y la esperada es: $E^*6 = 10,41+3,12+0,8+0,18+0,05= 14,56$) Luego de efectuar este arreglo, se calculan los valores de χ^2 respectivos que se colocan en la sexta columna y se obtiene un total de $\chi^2 = 7,24$. Si se aplica la corrección de Yates el valor disminuye mucho $\chi^2_{corr} = 6,515$.

Los grados de libertad se calculan de la manera siguiente: Hay 10 clases de frecuencias observadas, pero con el agrupamiento quedan $n^* = n-4 = 6$. En una distribución de Poisson se usa un solo dato poblacional μ , luego resulta $r = 1$. Entonces, los grados de libertad son:

$$v = n^* - r - 1 = (10 - 4) - 1 - 1 = 4$$

De tablas es $\chi^2_{0,95; 4} = 9,488 > \chi^2_{corr} = 6,515$ por lo tanto no se puede rechazar la hipótesis nula. En cambio, si se hubieran estimado las probabilidades con la función de Gauss en vez de la Poisson, no habría sido lo mismo: con Poisson se ajusta mejor que la de Gauss

Como los grados de libertad dependen del número de parámetros poblacionales que se usan con cada modelo teórico, y con el número de clases luego de ser agrupadas, las fórmulas para calcularlos más habituales son:

Tabla 16.2 Grados de libertad

Distribución	Parámetros pob.	r	$\nu = n^* - r - 1$	Donde n^* se calcula con el número de clases, del problema analizado, luego de hacerse la agrupación para que no ocurra que $n < 5$
Poisson	μ	1	$n^* - 2$	
Gauss	$\mu ; \sigma$	2	$n^* - 3$	
Binomial	p	1	$n^* - 2$	

16.2 El método moderno con G-test

Como se vio en el capítulo anterior, el G-test es más sensible y robusto que la prueba de la Chi cuadrado. Por lo tanto, se torna el método más recomendable para analizar el problema de la bondad de ajuste. Su uso es totalmente análogo al visto más arriba, con la salvedad de emplear otro estadígrafo G, para compararlo con los valores críticos de la Chi cuadrado y poder decidir acerca de la hipótesis nula. Aplicando este modelo a los datos de la Tabla 16.1 anterior se tiene:

Tabla 16.3 La prueba de G con Poisson.

N	O _i	E _i	O* _i	E* _i	G _i
0	75	66,12	75	66,12	18,903
1	103	119,02	103	119,02	-29,780
2	121	107,11	121	107,11	29,508
3	54	64,27	54	64,27	-18,804
4	30	28,92	30	28,92	2,200
5	13	10,41	13	10,41	5,777
6	2	3,12	4	4,15	-0,295
7	1	0,80			
8	0	0,18			
9 y más	1	0,05			
Total	400	400,00	400	400	7,509

Los grados de libertad en este caso son:

$$\nu = n^* - r - 1 = (10-3) - 1 - 1 = 5$$

Y la corrección de Williams es:

$$Q = 1 + [(k^2 - 1)] / 6N\nu$$

Con $k = 10$: número de clases
 $N = 400$: número de casillas

$$Q = 1 + [(100-1) / (6 \cdot 400 \cdot 5)] = 1,00825$$

Si la frecuencia esperada es muy chica ($E_i < 3$), los valores de G no se pueden calcular. Una regla práctica es agrupar las frecuencias anexas tal que E^*_{i} sea mayor o igual a 3. En la Tabla 16.2 como las tres últimas frecuencias esperadas son chicas, se agrupan con la más cercana (la sexta) y lo mismo se hace con las observadas: O^*_{i} . Entonces, los datos agrupados se muestran en las columnas cuarta y quinta de la tabla ($O^*_{7} = 3,12+0,8+0,18+0,05 = 4,15$ y la observada es: $O^*_{5} = 2+1+0+1 = 4$). Luego de efectuar este arreglo, se calculan los valores de G respectivos, que se colocan en la sexta columna y se obtiene un total de $G = 7,509$. Si se aplica la corrección de Williams resulta $G_{corr} = G / Q = 7,50825 / 1,00825 = 7,45$. Como los grados de libertad son $\nu = 4$ de es $\chi^2_{0,95; 4} = 9,488 > G_{corr} = 7,45$ por lo tanto no se puede rechazar la hipótesis nula.

En el tema 7 se desarrolló un ejemplo de aplicación usando la probabilidad binomial, para estudiar el caso de 6.115 familias con 12 hijos. En el cuadro 7.1 se volcaron los datos de las frecuencias observadas y calculadas. Tomando esa información se prepara la Tabla 16.4 para ilustrar el uso de la prueba G con una población supuesta como binomial:

Tabla 16.4 La prueba de G con la Binomial.

Hijas mujeres	Observadas Oi	Esperadas Ei	O*i	E*i	Gi
0	7	2,346			
1	45	26,082	52	28,428	62,802
2	181	132,836	181	132,836	111,996
3	478	410,016	478	410,016	146,664
4	829	854,247	829	854,247	-49,740
5	1112	1265,628	1112	1265,628	-287,804
6	1343	1367,279	1343	1367,279	-48,124
7	1033	1085,211	1033	1085,211	-101,869
8	670	628,055	670	628,055	86,631
9	286	258,475	286	258,475	57,882
10	104	71,803	104	71,803	77,057
11	24	12,089	27	13,022	39,377
12	3	0,933			
Total	6115	6115	6115	6115	94,872

Los grados de libertad son:
 $\nu = n^* - r - 1 = 9$

La corrección de Williams es:
 $Q = 1 + [(k^2 - 1)] / 6N\nu$

Con $k = 11$
 $N = 6.115$ resulta
 $Q = 1 + [(121-1)/(6 \cdot 6115 \cdot 9)]$
 $Q = 1,0003634$

$G_{corr} = 94,83^{***}$

$\chi^2_{0,999; 9} = 27,877 \lll G_{corr}$

Se rechaza la hipótesis nula que suponía una distribución binomial en la población. Se encontró evidencia altamente significativa que demuestra que el ajuste no es bueno.

En el tema 9 se presentó un ejemplo sobre las distribuciones en peso de 195 varones. Los datos mostrados en el Cuadro 9.4 se vuelcan a continuación en las dos primeras columnas de la Tabla 16.5, para poder usar el G-test en el caso de una distribución poblacional gaussiana:

Tabla 16.5 La prueba de G con la Normal.

Intervalos del peso	Observadas Oi	Esperadas Ei	O*i	E*i	Gi
hasta 63,5	3	3,3			
63,5-65,5	14	12,2	17	15,5	3,140693
65,5-67,5	30	31,7	30	31,7	-3,307158
67,5-69,5	48	51,1	48	51,1	-6,008015
69,5-71,5	48	50,7	48	50,7	-5,25359
71,5-73,5	39	31,1	39	31,1	17,6556
73,5-75,5	11	11,7	13	14,9	-3,546708
75,5 y más	2	3,2			
Total	195	195	6115	195	2,68082

Los grados de libertad son:
 $\nu = n^* - r - 1 = 6 - 2 - 1 = 3$

La corrección de Williams es:
 $Q = 1 + [(k^2 - 1)] / 6N\nu$

Con $k = 6$
 $N = 195$ resulta
 $Q = 1 + [(36-1)/(6 \cdot 195 \cdot 3)]$
 $Q = 1,00598$

$G_{corr} = 2.664 < \chi^2_{0,95;3} = 11,07$

Para este caso, al número total de clases útiles $k=6$ hay que restarle $r = 2$ grados de libertad pues se usan los datos muestrales para estimar a los dos parámetros poblacionales (μ ; σ) necesarios para obtener las probabilidades gaussianas. Efectuada la corrección de Williams, el valor G_{corr} es menor que el crítico de tablas para el 95% de confianza. Por lo tanto, no se puede rechazar la hipótesis de que los pesos tienen una distribución normal en la población.

16.3 La prueba de Kolgomorov-Smirnov

Se trata de un modelo *no-paramétrico* para comprobar bondad de ajuste. La clasificación de este modelo se debe al hecho que no depende de una distribución original específica; no se sabe a priori la forma de la distribución de donde se sacan los datos. Por ejemplo el modelo de Student requiere una población original de tipo gaussiana, como la de Chi-cuadrado y la de Fisher. Aquí se trata de *distribuciones libres*, sin parámetros de tipo prefijado como μ o σ . Es cuando se está más interesado en comparar distribuciones antes que parámetros, ni tampoco cuando se trata de estimar parámetros (*modelos paramétricos*). El modelo de Kolgomorov-Smirnov se aplica a distribuciones de tipo continuas y es considerada *conservadora*. También se la usa para probar hipótesis acerca de distribuciones discretas. Es un modelo fácil de realizar. Se basa en calcular las diferencias, en valor absoluto, entre las frecuencias acumuladas relativas observadas y las esperadas, en cada clase. Luego se busca la mayor de las diferencias en valor absoluto, y el estadígrafo $D_{\text{máx}}$ así obtenido se compara con el valor crítico de tablas. En este punto de trata del modelo aplicado al caso de una sola muestra y la Tabla 13 del anexo se presentan los valores críticos para validar hipótesis. Este modelo presenta especiales ventajas cuando se lo aplica a muestras pequeñas, y en tales casos no se necesita agrupar en clases. Otra ventaja es que permite construir límites de confianza para la distribución acumulada completa.

$$D_{\text{máx}} = | O_{Ai} - E_{Ai} |_{\text{máx}}$$

Donde O_{Ai} : frecuencia acumulada relativa de la clase i observada
 E_{Ai} : frecuencia acumulada relativa de la clase i esperada

Para comparar este estadígrafo con un valor crítico de tablas, se busca ese valor en tablas entrando con el tamaño muestral N . Hay dos tablas para los valores críticos de esta distribución. La Tabla 13 que se emplea para los casos más comunes donde se usan los datos muestrales para estimar los valores poblacionales, o sea, para cuando se usan *hipótesis intrínsecas*. A su vez, la Tabla 14 es también para Kolgomorov-Smirnov aplicado a una sola muestra, pero para los raros casos donde se usan *hipótesis extrínsecas*.

La Tabla 13 se extiende correlativamente hasta $N = 30$, y luego a partir de allí se usa una aproximación dada por las relaciones siguientes:

Cuadro 16.1 Valores críticos para la prueba de Kolgomorov-Smirnov en una sola muestra.

Hipótesis intrínsecas			
$N > 30$	α	$K\alpha$	$D\alpha = K\alpha / \sqrt{N}$
	0,10	0,805	$0,805 / \sqrt{N}$
	0,05	0,886	$0,886 / \sqrt{N}$
	0,01	1,031	$1,031 / \sqrt{N}$

Significativo
Muy significativo

La Tabla 14 del Anexo es hasta una muestra de tamaño 100. Por lo tanto se debe usar la aproximación asintótica dada por:

$$D\alpha = \sqrt{\frac{-\ln(\alpha/2)}{2N}} = K\alpha / \sqrt{N}$$

Y para los valores críticos usuales se puede preparar otro cuadro resumen:

Cuadro 16.2 Valores críticos para la prueba de Kolgomorov-Smirnov en una sola muestra. Hipótesis extrínsecas

N > 100	α	$K\alpha$	$D\alpha = K\alpha / \sqrt{N}$	
	0,05	1,358	$1,358 / \sqrt{N}$	Significativo
	0,01	1,628	$1,628 / \sqrt{N}$	Muy significativo
	0,001	1,949	$1,949 / \sqrt{N}$	Altamente significativo

Este modelo en muchos casos posee mayor potencia que la prueba de la Chi cuadrado para bondad de ajuste. Su uso es más sencillo y es la recomendada para realizar este tipo de estudios en el área de los análisis clínicos y farmacológicos. Solo es aventajada en sensibilidad por el G-test, pero se puede aplicar a casi todos los casos. Para mostrar el uso de esta prueba se desarrollan los ejemplos siguientes, usando los datos de los casos vistos en el punto anterior:

Ejemplo 1) Aplicar el test al caso visto en la Tabla 16.3 de informe de Student en 1907.

n	Oi	Ei	Oacum	Eacum	OAi	EAi	OAi - EAi	
0	75	66,12	75	66,12	0,1875	0,1653	0,0222	→ Dmáx
1	103	119,02	178	185,14	0,4450	0,4629	0,0179	
2	121	107,11	299	292,25	0,7475	0,7306	0,0169	
3	54	64,27	353	356,52	0,8825	0,8913	0,0088	
4	30	28,92	383	385,44	0,9575	0,9636	0,0061	
5	13	10,41	396	395,85	0,9900	0,9896	0,0004	
6	2	3,12	398	398,97	0,9950	0,9974	0,0024	
7	1	0,80	399	399,77	0,9975	0,9994	0,0019	
8	0	0,18	399	399,95	0,9975	0,9999	0,0024	
9 y más	1	0,05	400	400	1,0000	1,0000	0,0000	
Total	400	400,00						

Para este problema como se usan los datos muestrales para estimar μ con la media y aplicar Poisson se trata de un caso de hipótesis intrínsecas. Luego será:

$$D\alpha = K\alpha / \sqrt{N} = 0,886 / 20 = 0,0443 > Dmáx = 0,0222 \quad (95\%)$$

Y no se puede rechazar la hipótesis nula. La misma conclusión que la obtenida con el G-test. Aunque el resultado está muy cerca del límite para una confianza del 95%, alcanza para ver la aplicación de este modelo menos poderoso que el G-test.

Ejemplo 2) Aplicar el test al caso visto en la Tabla 16.4, del estudio de 6115 familias de Sajonia con 12 hijos, y ver la distribución de sexos entre ellos; n es la cantidad de hijas mujeres. Se postula una distribución binomial en esta población

n	Oi	Ei	Oacum	Eacum	OAi	EAI	OAI- EAI
0	7	2,346	7	2,346	0,0011	0,0004	0,0007
1	45	26,082	52	28,428	0,0085	0,0046	0,0039
2	181	132,84	233	161,26	0,0381	0,0264	0,0117
3	478	410,02	711	571,28	0,1163	0,0934	0,0228
4	829	854,25	1540	1425,5	0,2518	0,2331	0,0187
5	1112	1265,6	2652	2691,2	0,4337	0,4401	0,0064
6	1343	1367,3	3995	4058,4	0,6533	0,6637	0,0104
7	1033	1085,2	5028	5143,6	0,8222	0,8412	0,0189
8	670	628,06	5698	5771,7	0,9318	0,9439	0,0121
9	286	258,48	5984	6030,2	0,9786	0,9861	0,0076
10	104	71,803	6088	6102	0,9956	0,9979	0,0023
11	24	12,089	6112	6114,1	0,9995	0,9998	0,0003
12	3	0,933	6115	6115	1	1,0000	0,0000
Total	6115	6115					

→ D_{máx}

El valor crítico para hipótesis intrínsecas es:

$$D\alpha = K\alpha / \sqrt{N} = 0,886 / 78,2 = 0,0113 < D_{máx} = 0,0228 \text{ (95\%)}$$

Y

$$D\alpha = K\alpha / \sqrt{N} = 1,031 / 78,2 = 0,0132 < D_{máx} = 0,0228 \text{ (99\%)}$$

Por lo tanto se tiene evidencia muy significativa para rechazar la hipótesis de una distribución binomial en la población.

Ejemplo 2) Aplicar el test al caso visto en la Tabla 16.5, del estudio del peso de 195 varones suponiendo una distribución Normal del peso en la población.

Peso	Oi	Ei	Oacum	Eacum	OAI	EAI	OAI-EAI
hasta 63,5	3	3,3	3	3,3	0,0154	0,0169	0,0015
63,5-65,5	14	12,2	17	15,5	0,0872	0,0795	0,0077
65,5-67,5	30	31,7	47	47,2	0,2410	0,2421	0,0010
67,5-69,5	48	51,1	95	98,3	0,4872	0,5041	0,0169
69,5-71,5	48	50,7	143	149,0	0,7333	0,7641	0,0308
71,5-73,5	39	31,1	182	180,1	0,9333	0,9236	0,0097
73,5-75,5	11	11,7	193	191,8	0,9897	0,9836	0,0062
75,5 y más	2	3,2	195	195,0	1,0000	1,0000	0,0000
Total	195	195					

→ D_{máx}

El valor crítico para hipótesis intrínsecas es:

$$D\alpha = K\alpha / \sqrt{N} = 0,886 / 13,96 = 0,0635 > D_{\text{máx}} = 0,0308 \text{ (95\%)}$$

Y por lo tanto no se puede rechazar la hipótesis nula, de una distribución gaussiana de los pesos.

16.4 Test de Kolgomorov-Smirnov para 2 muestras

Es un modelo estadístico para verificar si *dos muestras independientes* han sido extraídas de la misma población o de poblacionales con igual función distribución. La prueba de dos colas es sensible a cualquier clase de diferencia entre las distribuciones de donde fueron extraídas ambas muestras. La prueba de una cola se usa para decidir si los valores de la población, de donde se extrajo a una de las muestras, son estocásticamente mayores (o menores) que las de la otra población. Por ejemplo, para probar la predicción de que los valores de un grupo experimental serán mejores que los del grupo control.

Como en el caso de una sola muestra, el modelo compara las dos frecuencias acumulativas relativas de ambas muestras y busca la mayor diferencia encontrada. Este valor multiplicado por ambos tamaños muestrales, se contrasta contra un valor de tablas. Sean las muestras A y B, con un tamaño muestral de N_A y N_B respectivamente, y con $D_{\text{máx}}$ como la mayor diferencia encontrada en valor absoluto, entonces el valor de K se calcula con:

$$K = N_A \cdot N_B \cdot D_{\text{máx}}$$

Si este valor K es mayor que el de tablas $K\alpha$ para un nivel de significación dado, se rechaza la hipótesis nula de que ambas muestras provienen de la misma población. Cuando el tamaño de muestras es menor a 25 se pueden usar las Tablas 14 del Anexo. Cuando las muestras son grandes se puede emplear una aproximación, como se muestra al final de las Tablas 14, donde se da el valor crítico de $D\alpha$ para comparar con $D_{\text{máx}}$. Finalmente, en la Tabla 15 y 16 del Anexo se presenta una versión simplificada de las tablas, para el caso de igual tamaño muestral, o sea cuando $N_A = N_B = n$; en esta última también se expresan los valores de $D_{\text{máx}}$ a comparar con $K\alpha$. Para ilustrar estas ideas se presenta el caso siguiente:

Ejemplo 1) De un estudio publicado por la *J. Amer. Med. Ass.* De Friedman, M. Et al., 217, pág. 929-932 (1971) se sacaron los datos de la tabla siguiente:

N	1	2	3	4	5	6	7	8	9	10	11
Grupo A	3,6	2,6	4,7	8,0	3,1	8,8	4,6	5,8	4,0	4,6	
Grupo B	16,2	17,4	8,5	15,6	5,4	9,8	14,9	16,6	15,9	5,3	10,5

A dos grupos de personas, descansados y luego de ingerir una infusión de Hipoclorito de argenina, se le midió los niveles pico de la hormona de crecimiento en plasma. Los sujetos se clasificaron en dos grupos. Grupo A: se trata de individuos relativamente predispuestos a enfermedades coronarias, cuyo temperamento se caracteriza por un excesivo sentido de competitividad, lide-

razgo y urgencia de tiempo. Grupo B: estos son relativamente resistentes a tener problemas coronarios y cuyo temperamento es lo inverso del grupo anterior. Estudios anteriores del mismo equipo (1950) indicaban que el Grupo A tenía mayor tendencia a sufrir enfermedades coronarias que los del Grupo B. En este trabajo, se investiga la incidencia del temperamento en la cantidad de hormona de crecimiento estimulada con la argenina.

La hipótesis de trabajo es que ambos grupos deben diferir. O sea, ambos grupos no provienen de la misma población. Se plantea entonces como

$H_0 : P (X \leq a) \geq P (Y \leq a)$ No hay diferencia entre ambas muestras.

$H_1 : P (X \leq a) < P (Y \leq a)$ El nivel hormonal es menor en el Grupo A que en el B.

Para desarrollar este modelo se deben seguir los pasos siguientes:

Paso 1) Se ordena cada grupo en orden ascendente. Con el rango se obtiene el número de clases, como en los histogramas, pero tratando de tener el mayor número de clases posible, para ambas muestras. Los resultados se muestran en la primer columna del cuadro siguiente.

Paso 2) Se calculan las frecuencias de cada grupo en las mismas clases, luego sus frecuencias acumuladas y finalmente, dividiendo por el respectivo tamaño muestral, se calculan las frecuencias acumuladas relativas, que se muestran en las columnas 2 y 3.

Paso 3) Para cada clase, se calculan las diferencias en valor absoluto de las frecuencias acumuladas relativas de cada grupo. Y se busca el valor máximo de esas diferencias $D_{\max} = 0,7$

Paso 4) Se determina cual tabla de valores críticos usar. En este caso, se trata de muestras pequeñas con diferente tamaño muestral (Tabla 15) Con $N_A = 10$ y $N_B = 11$ se obtiene de tablas los valores críticos: $K_{0,05} = 60$; $K_{0,01} = 77$ y $K_{0,001} = 89$

Paso 5) Se comparan estos valores críticos con el estadígrafo K de Kolgomorov-Smirnov para dos muestras: $N_A \cdot N_B \cdot D_{\max} = K = 10 \cdot 11 \cdot 0,7 = 77$

	Grupo A	Grupo B	Diferencias
2,5 - 3,4	0,2	0	0,200
3,5 - 4,4	0,4	0	0,400
4,5 - 5,4	0,7	0	0,700 (D_{máx})
5,5 - 6,4	0,8	0,182	0,618
6,5 - 7,4	0,8	0,182	0,618
7,5 - 8,4	0,9	0,273	0,627
8,5 - 9,4	1	0,364	0,636
9,5 - 10,4	1	0,455	0,545
10,5 - 11,4	1	0,455	0,545
11,5 - 12,4	1	0,455	0,545
12,5 - 13,4	1	0,455	0,545
13,5 - 14,4	1	0,455	0,545
14,5 - 15,4	1	0,545	0,455
15,5 - 16,4	1	0,818	0,182
16,5 - 17,4	1	1	0,000

Se concluye que se debe rechazar la hipótesis nula y aceptar la alternativa. O sea, se tiene evidencia muy significativa $K = 77^{**}$ de que ambas muestras no provienen de la misma población, los valores hormonales del Grupo A son menores que los del Grupo B.

16.5 Test de Bondad de ajuste con repetición

Es un modelo estadístico para los casos en que los datos para un test se hacen de manera repetida. La manera más simple es agrupar los datos repetidos en una especie de “pool” de datos y usar los totales para hacer la prueba. Si bien se respeta la regla de oro de maximizar los datos, el problema es que se puede perder información a causa del agrupamiento y así no tener un panorama mejor. Para mostrar un caso de este tipo se presenta el ejemplo siguiente:

Ejemplo 1) Una industria farmacéutica encara un estudio sobre el efecto de un nuevo analgésico. Para ello, efectúa relevamiento en 8 ciudades diferentes de un país elegidas al azar, donde se efectúa la misma prueba. Todos los casos analizados con respuesta positiva, se dividen según el sexo. La hipótesis de trabajo es que este factor no tiene influencia en los resultados obtenidos. Los datos obtenidos se muestran en la tabla siguiente. Con los cuales se realiza un G-test como se vio en el punto 15.2 del capítulo anterior:

Sexo	Observadas	Esperadas	G	Como $\chi^2_{0,999; 1} = 10,828$ los resultados fueron altamente significativos y se tiene una muy fuerte evidencia para rechazar H_0
Femenino	616	530	185,26	
Masculino	444	530	-157,22	
Total	1060	1060	28,04***	

El farmacéutico a cargo del estudio estadístico deber rechazar la hipótesis del 50% para cada sexo, parecería que este factor tiene algo que ver. Sin embargo, antes de tomar una decisión al respecto, decide aprovechar al máximo los datos recogidos, con el desglose en las 8 localidades y enfoca el estudio desde otro ángulo:

Caso 1) Test de independencia (tabla de contingencia)

Frecuencias Observadas

Loc.	f	m	Total
1	83	47	130
2	77	43	120
3	110	96	206
4	92	58	150
5	51	31	82
6	48	61	109
7	70	42	112
8	85	66	151
Total	616	444	1060

Frecuencias Esperadas

Loc.	f	m	Total
1	75,547	54,453	130
2	69,736	50,264	120
3	119,71	86,287	206
4	87,17	62,83	150
5	47,653	34,347	82
6	63,343	45,657	109
7	65,087	46,913	112
8	87,751	63,249	151
Total	616	444	1060

Coloca los datos observado en una tabla como la de arriba y para el cálculo de las frecuencias observadas, usa el concepto de independencia. Es decir, multiplicando los totales marginales de cada casilla y dividiendo por $N = 1060$. En la primer casilla será $E_{11} = (130 \cdot 616) / 1060 = 75,547$ y así sucesivamente completa la tabla de frecuencias esperadas de más arriba. Luego hace el test:

Loc.	Gf	Gm	Gf+Gm
1	15,61784	-13,8356	1,782242
2	15,26	-13,4239	1,836087
3	-18,6161	20,48096	1,864869
4	9,923213	-9,27915	0,644065
5	6,924128	-6,35701	0,567118
6	-26,6275	35,34652	8,719038
7	10,1883	-9,29289	0,895414
8	-5,41473	5,619838	0,205107
Tot	7,25518	9,25876	16,51394

G = 16,51394

Para cada casilla se calcula el valor:

$$G_{ij} = 2 O_{ij} \cdot \ln(O_{ij} / E_{ij})$$

La suma total para filas y columnas es:

$$G = 16,514 *$$

Los grados de libertad son: $\nu = (8-1)(2-1) = 7$

De tablas $\chi^2_{0,95; 7} = 14,067$

Se tiene prueba (95%) de *heterogeneidad* en los datos. Esto significa, que el factor Sexo, no es independiente del factor Localidad. En la tercer columna se coloca el valor Gf + Gm para poder realizar otro tipo de análisis. Comparando el peso que tiene cada localidad en el total, se nota que más del 50% del valor total de G se debe a la Localidad 6, mientras que las demás no parecen ser significativas, porque tienen una contribución parecida en el total. El investigador ahora sospecha que los resultados inesperados del primer análisis que efectuó, se originan en la gran desproporción encontrada en esta localidad. Cosa que deberá ser investigada con mucho cuidado.

En realidad, con cada localidad se puede armar una tabla 2x2, usando los datos observados en ella, y para obtener los esperados se usa una estimación del porcentaje poblacional. Esto es, la probabilidad de que sea de sexo femenino es $P(f) = 616 / 1060 = 0,581132$ y para el masculino es: $P(m) = 444 / 1060 = 0,418868$. Entonces, en una localidad cualquiera, por ejemplo la 6, se tendrá una frecuencia esperada femenina $E_f = 109 \cdot 0,581132 = 63,34$ y masculina $E_m = 109 \cdot 0,418868 = 45,66$ lo que se obtiene multiplicando la probabilidad por el tamaño muestral de la localidad. Se puede entonces armar una tabla para esa localidad como la siguiente :

Sexo	Observadas	Esperadas	G	Como $\chi^2_{0,95; 1} = 3,841$ los resultados fueron a significativos y se tiene evidencia para rechazar H_0
Femenino	48	63,34	-26,628	
Masculino	61	45,66	35,347	
Total	109	109	8,719	

Realizado el test de G, resultó significativo pues $G = 8,719 * > \chi^2_{0,95; 1} = 3,841$. Análogamente se pueden plantear los G-test por localidad y comparar siempre contra el mismo valor crítico. Eso se muestra en la última columna de la tabla al principio de esta página. Se concluye que la única localidad con diferencia significativa es la Localidad 6. Como esto no es la tendencia general, algo debe haber ocurrido en tal localidad para tener esos resultados. Algo que merece ser investigado con cuidado en un próximo experimento, que deberá ser planeado cuidadosamente.

Se puede hacer un resumen de estos resultados calculando G agrupado = 1,5004 que se calcula usando $O_i = 616$; $O_j = 444$ y $E_i = 1060 (9/16)$; $E_j = 1060 (7/16)$

Tests realizados	Grados de libertad	Estadígrafo G
Total Datos Agrupados	1	1,500
Total por localidad	7	16,514*
TOTAL	8	18,014*

16.6 Problemas propuestos

1) Marcar la respuesta correcta a cada una de las afirmaciones siguientes, o completar la frase:

- | | | |
|--|----------|----------|
| 1) Bondad de ajuste significa que tan bien se parecen los datos reales a los teóricos. | V | F |
| 2) El método clásico para efectuar una prueba de bondad de ajuste es | | |
| 3) Se acostumbra a agrupar las clases con sus anexas cuando la frecuencia observada es < 3 . | V | F |
| 4) Para el cálculo de los grados de libertad se usa el número de clases agrupadas. | V | F |
| 5) El G-test comparado con el clásico es mejor porque es | | |
| 6) Cuando la frecuencia esperada en un G-test es menor que 5 hay que agrupar las clases. | V | F |
| 7) El método de Kolgomorov-Smirnov (KS-test) busca la diferencia máxima de frecuencias. | V | F |
| 8) El KS-test se puede usar para una sola muestra o para comparar dos entre sí. | V | F |
| 9) Los pasos a seguir en un KS-test para una sola muestra son | | |
| 10) Da lo mismo usar hipótesis extrínsecas que intrínseca en un KS-test. | V | F |
| 11) El KS-test para dos muestras se puede usar si estas son independientes. | V | F |
| 12) Las tablas estadísticas para el KS-test de una y dos muestras son iguales. | V | F |
| 13) Los pasos a seguir para efectuar un KS-test en dos muestras son..... | | |
| 14) El estadígrafo de comparación en el KS-test es el mismo. | V | F |
| 15) Conviene ordenar los datos en forma creciente antes de usar el KS-test para 2 muestras. | V | F |
| 16) Un test de bondad de ajuste con repetición solo permite comparar los datos agrupados. | V | F |
| 17) Las pruebas que se pueden plantear en un test con repetición son | | |
| 18) La variabilidad total en un test de bondad de ajuste se cuantifica con..... | | |
| 19) Heterogeneidad de los datos repetidos implica encontrar significación en el G-test . | V | F |
| 20) El G de datos agrupados, más el G de datos desagrupados es igual al G total. | V | F |

2) Aplicar el KS-test a los tres problemas vistos en el punto 16.1 y comparar las conclusiones obtenidas con las del texto usando el G-test.

3) Ídem anterior para los problemas vistos en el punto 9.7 del capítulo 9.

4) Se efectuaron 1000 pruebas de efectividad de un nuevo medicamento, en 5 grupos de 200 individuos enfermos cada uno, elegidos al azar con el método de doble ciego. Los resultados se clasificaron en SI cuando fueron efectivos, y NO en caso contrario. Con esta información decidir si el medicamento es efectivo.

Caso	SI	NO	Total
1	120	80	200
2	110	90	200
3	98	102	200
4	130	70	200
5	125	75	200
Total	583	417	1000