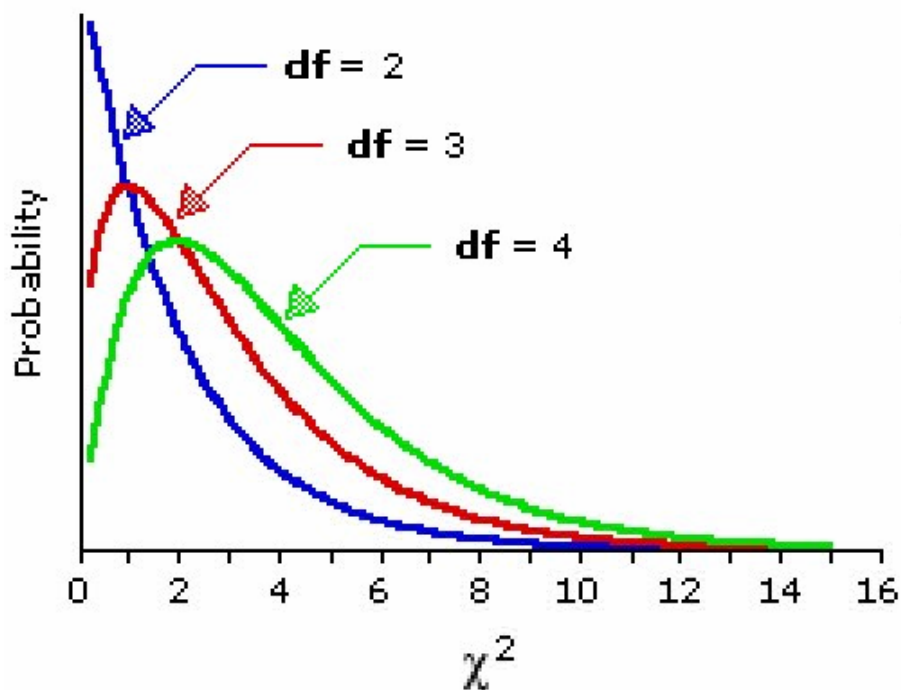


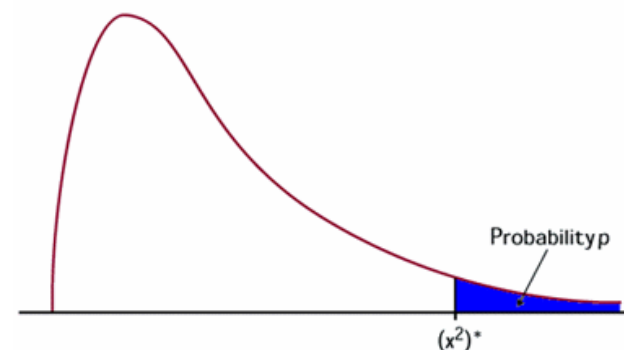


## 8.1. Los datos de frecuencias:

- Muy comunes (ejemplo)
- Difíciles de analizar por métodos paramétricos (no recomendable)
- Desarrollo de métodos propios:
  - Aproximación Chi-cuadrado ( $\chi^2$ )
  - Test G (máxima verosimilitud)
  - Otros (test exacto de Fisher, Monte Carlo, etc)
- El uso de la distribución  $\chi^2$  (distinto al test anterior)



	DATOS HOSPITAL X	
	Médicos	Enfermeros
Varón	34	56
Hembra	12	135





## 8.2. Los tests de bondad de ajuste:

- El objetivo es verificar una hipótesis propia (ajena a nuestros datos)
- Ejemplo datos de otros hospitales sugieren que la proporción de Varón/hembra en médicos es de 3 a 1.
- Estimación de los esperados bajo  $H_0$  (los observados son una muestra de los esperados)
- Cuantificación de las diferencias con observadas

	DATOS HOSPITAL X	
	Médicos	Enfermeros
Varón	34	56
Hembra	12	135

	DATOS Médicos	
	Varón	Hembra
Obs	34	12
Esp(3/1)	$\frac{3}{4}$	$\frac{1}{4}$

Esp	34,5	11,5
-----	------	------

$\chi^2$	0,01	0,02
----------	------	------

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$O$  = the frequencies observed

$E$  = the frequencies expected.

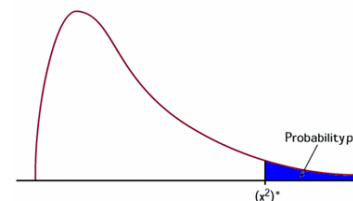
$\sum$  = the 'sum of'

$$\chi^2 = 0,01 + 0,02 = 0,03$$



## 8.2. Los tests de bondad de ajuste:

- Uso de la distribución  $\chi^2$ : test de hipótesis
- Los grados de libertad ( $n^0$  clases –  $n^0$  de parámetros)...



DATOS HOSPITAL X		
	Médicos	Enfermeros
Varón	34	56
Hembra	12	135

DATOS Médicos		
	Varón	Hembra
Obs	34	12
(Esp)	(34,5)	(11,5)

$\chi^2 = 0,03$

ya que  $\chi^2$  obtenido es menor que 3,841, la probabilidad de la desviación observada es alta ( $p > 0,05$ )

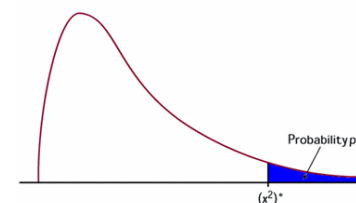
no podemos rechazar  $H_0$

df	upper tail area	0.99	0.95	0.90	0.10	0.05	0.01
1		0.00016	0.0039	0.016	2.706	<b>3.841</b>	6.635
2		0.020	0.103	0.211	4.605	5.991	9.210
3		0.115	0.352	0.584	6.251	7.815	11.34
4		0.297	0.711	1.064	7.779	9.488	13.28
5		0.554	1.145	1.610	9.236	11.07	15.09
10		2.558	3.940	4.865	15.99	18.31	23.21
15		5.229	7.261	8.547	22.31	25.00	30.58
20		8.260	10.85	12.44	28.41	31.41	37.57
25		11.52	14.61	16.47	34.38	37.65	44.31



## 8.2. Los tests de bondad de ajuste:

- Otro ejemplo



DATOS HOSPITAL X		
	Médicos	Enfermeros
Varón	34	56
Hembra	12	135

DATOS Médicos		
	Varón	Hembra
Obs	56	135
(Esp)	(143,2)	(47,8)

$$\chi^2 = 212,2$$

ya que  $\chi^2$  obtenido es mayor que 3,841, la probabilidad de la desviación observada es muy baja ( $p < 0,001$ )

podemos rechazar  $H_0$

df	upper tail area	0.99	0.95	0.90	0.10	0.05	0.01
1		0.00016	0.0039	0.016	2.706	<b>3.841</b>	6.635
2		0.020	0.103	0.211	4.605	5.991	9.210
3		0.115	0.352	0.584	6.251	7.815	11.34
4		0.297	0.711	1.064	7.779	9.488	13.28
5		0.554	1.145	1.610	9.236	11.07	15.09
10		2.558	3.940	4.865	15.99	18.31	23.21
15		5.229	7.261	8.547	22.31	25.00	30.58
20		8.260	10.85	12.44	28.41	31.41	37.57
25		11.52	14.61	16.47	34.38	37.65	44.31



## 8.2. Los tests de bondad de ajuste:

- Se desaconseja con alta proporción de celdas con esperados menores que 1
- Una variante conservadora es el uso de la corrección de Yates:
  - . Se disminuyen las diferencias entre Esp y Obs en 0,5
  - . Ejemplo

### SIN corrección

	DATOS Médicos	
	Varón	Hembra
Obs	56	135
(Esp)	(143,2)	(47,8)

$$X^2 = 212,2$$

### CON corrección

	DATOS Médicos	
	Varón	Hembra
Obs	56,5	134,5
(Esp)	(142,7)	(48,3)

$$X^2 = 205,9$$

en este caso las conclusiones no cambian,  
pero en otros más ajustados si podrían cambiar



## 8.2. Los tests de bondad de ajuste:

- El test G:

- . Sigue una distribución  $\chi^2$  mejor que el test  $\chi^2$
- . Ejemplo
- . También se puede aplicar la corrección de Yates ó de Williams

DATOS Médicos		
	Varón	Hembra
Obs	56	135
(Esp)	(143,2)	(47,8)

$$X^2 = 212,2$$

$$G = 2 \sum O_i \ln(O_i/E_i)$$

$$O \ln(O/E) \quad -52,6 \quad 140,16$$

$$G = 87,6$$

Corrección de Williams:

$$G_{adj} = G/q$$

podemos rechazar  $H_0$

$$G_{adj} = 87,3$$

$$q = 1 + \frac{(a^2 - 1)}{(6n)(gl)} = 1 + \frac{3}{(1146)(1)} = 1,003$$



## 8.3. Los tests de homogeneidad y heterogeneidad:

- La evaluación de homogeneidad respecto a un nivel de replicación determinado
- Para más de dos muestras
- Hipótesis dependiente de los datos (gl)
- Ejemplo (frecuencias semejantes entre hospitales; homogéneas):
  - . Estimar  $p$ (por sexo)
  - . Estimar  $p$ (por hospital)
  - .  $H_0 = p(\text{sexo})p(\text{hospital})$

Probabilidad por Sexo  
 $V = 99/130 = 0,762$   
 $H = 31/130 = 0,238$

DATOS Médicos			
	Varón	Hembra	Tot
H1	34 $E_1$	12	46
H2	40	7	47
H3	25	12	37
Tot	99	31	130

Probabilidad por Hospital  
 $H_1 = 46/130 = 0,354$   
 $H_2 = 47/130 = 0,362$   
 $H_3 = 37/130 = 0,284$

	V	H
H1	35	11
H2	35,8	11,2
H3	28,2	8,8

$E_1 = \text{TOT} \times V \times H_1 = 130 \times 0,762 \times 0,354 = 35,0$

$E_1 = (46 \times 99)/130$



## 8.3. Los tests de homogeneidad:

- Ejemplo (frecuencias semejantes entre hospitales; homogéneas):
  - . Obtener el valor de  $\chi^2$
  - . Obtener los grados de libertad ( $f-1 \times c-1 = n - p$ )
  - . Toma de decisiones

	Varón	Hembra
H <sub>1</sub>	34 (35)	12 (11)
H <sub>2</sub>	40 (35,8)	7 (11,2)
H <sub>3</sub>	25 (28,2)	12 (8,8)

	valor de $\chi^2$	
H <sub>1</sub>	0,03	0,09
H <sub>2</sub>	0,49	1,57
H <sub>3</sub>	0,36	1,16

$$\chi^2 = 3,7$$

$$gl = (3-1) \times (2-1) = 2$$

La probabilidad de 3,7 es mayor de 0,05

No se rechaza la H<sub>0</sub> de homogeneidad





## 8.3. Los tests de homogeneidad:

- El test G de homogeneidad:
  - . Sigue una distribución  $\chi^2$
  - . Ejemplo (frecuencias semejantes entre hospitales; homogéneas):

DATOS Médicos			
	Varón	Hembra	Tot
H <sub>1</sub>	34	12	46
H <sub>2</sub>	40	7	47
H <sub>3</sub>	25	12	37
Tot	99	31	130

$$G_H = 2(G_{\text{cel}} - G_{\text{col}} - G_{\text{fil}} + G_{\text{tot}})$$

$$\left. \begin{array}{l} G_{\text{cel}} \\ G_{\text{col}} \\ G_{\text{fil}} \\ G_{\text{tot}} \end{array} \right\} \sum O_i \ln O_i$$

$$\begin{aligned} G_{\text{cel}} &= 34\ln 34 + 12\ln 12 + 40\ln 40 + 7\ln 7 + 25\ln 25 + 12\ln 12 = 421,18 \\ G_{\text{col}} &= 99\ln 99 + 31\ln 31 = 561,37 \\ G_{\text{fil}} &= 46\ln 46 + 47\ln 47 + 37\ln 37 = 490,68 \\ G_{\text{tot}} &= 130\ln 130 = 632,78 \end{aligned}$$

$$G_H = 2(421,18 - 561,37 - 490,68 + 632,78) = 3,82$$

con  $(f-1) \times (c-1)$  gl (es decir 2), el valor no es significativo



## 8.4. Los tests de asociación/contingencia:

- El objetivo es diferente que el test de homogeneidad
- La calculística es idéntica al de homogeneidad ( $\chi^2$  y G)
- También se pueden aplicar correcciones de Yates o Williams

		DATOS HOSPITAL 1		
		Médicos	Enfermeros	
Varón	34	56	90	
Hembra	12	135	147	
	46	191	237	

$$G_I = 2(G_{\text{cel}} - G_{\text{col}} - G_{\text{fil}} + G_{\text{tot}})$$

claramente las proporciones de sexos difieren entre profesiones

$$\begin{aligned} G_{\text{cel}} &= 34\ln 34 + 12\ln 12 + 56\ln 56 + 135\ln 135 = 1037,35 \\ G_{\text{col}} &= 90\ln 90 + 147\ln 147 = 1138,58 \\ G_{\text{fil}} &= 46\ln 46 + 191\ln 191 = 1003,18 \\ G_{\text{tot}} &= 237\ln 237 = 1295,93 \end{aligned}$$

$$G_H = 2(1037,35 - 1138,58 - 1003,18 + 1295,93) = 383,03$$

con  $(f-1) \times (c-1)$  gl (es decir 1), el valor es altamente significativo



8.4. Test de bondad de ajuste múltiple (con replicación):

- $H_0$  = proporción laboral de hombres/mujeres 2/1
- Evaluación de  $H_0$  en 3 hospitales:
  - Se puede descomponer aditivamente
  - $G_{TOTAL} = G_{H1} + G_{H2} + G_{H3}$
  - $G_{TOTAL} = G_{AGR} + G_{Het}$

$$G = 2 \sum O_i \ln(O_i/E_i)$$

$$G_{HET} = G_{TOT} - G_{AGR}$$

$$G_{Het} = 2(G_{cel} - G_{col} - G_{fil} + G_{tot})$$

Observados

	Varón	Hembra
H <sub>1</sub>	34	12
H <sub>2</sub>	40	7
H <sub>3</sub>	25	12
AGR.	99	31

Esperados

30,7	15,3
31,3	15,7
24,7	12,3
86,7	43,3

G de bondad

GL

$G_{H1} = 2(34\ln(34/30,7) + 12\ln(12/15,3)) =$	1,12 <sup>ns</sup>	1
$G_{H2} = 2(40\ln(40/31,3) + 7\ln(7/15,7)) =$	8,22 <sup>**</sup>	1
$G_{H3} = 2(25\ln(25/24,7) + 12\ln(12/12,3)) =$	0,02 <sup>ns</sup>	1
$G_{Agr} = 2(99\ln(99/86,7) + 31\ln(31/43,3)) =$	5,54 <sup>*</sup>	1

$$G_{TOTAL} = G_{H1} + G_{H2} + G_{H3} = 9,36^*$$

$$GL = 3$$

$$G_{HET} = 9,36 - 5,54 = 3,82^{ns}$$

$$GL = 3 - 1 = 2$$

$$G_{HET} = 2(421,2 - 490,7 - 561,4 + 632,8) = 3,82^{ns}$$

G de heterogeneidad

$G_{cel} = 34\ln34 + 12\ln12 + 40\ln40 + 7\ln7 + 25\ln25 + 12\ln12 =$	421,18
$G_{col} = 46\ln46 + 47\ln47 + 37\ln37 =$	490,68
$G_{fil} = 99\ln99 + 31\ln31 =$	561,37
$G_{Tot} = 130\ln130 =$	632,78



# Referencias Bibliográficas

## LIBROS:

Sokal, R.R., Rohlf, F.J. 1995. Biometry. Freeman and co., New York

## PÁGINA WEB:

<http://udel.edu/~mcdonald/statgtestgof.html>

(Página Web donde se explica el test G y otros tests relacionados)