

# Estadística II

## Tema 2: Contrastes no paramétricos

José R. Berrendero

Departamento de Matemáticas  
Universidad Autónoma de Madrid

## Estructura de este tema

- ▶ Contrastes de bondad de ajuste
  - ▶ Contrastes basados en la distribución  $\chi^2$ .
  - ▶ Contraste de Kolmogorov-Smirnov.
  - ▶ Contrastes específicos para la distribución normal.
- ▶ Contrastes de homogeneidad e independencia basados en la distribución  $\chi^2$ .

## Contrastes no paramétricos

**Hipótesis no paramétricas:** no se pueden escribir en función de un número finito de parámetros.

- Bondad de ajuste:** A partir de una muestra  $X_1, \dots, X_n \equiv F$  de variables aleatorias iid, contrastar:
  - ▶  $H_0 : F = F_0$  donde  $F_0$  es una distribución prefijada.
  - ▶  $H_0 : F \in \{F_\theta : \theta \in \Theta\}$ , donde  $\Theta$  es el espacio paramétrico.
- Homogeneidad:** Dados  $X_1, \dots, X_n \equiv F$  y  $Y_1, \dots, Y_n \equiv G$  de variables aleatorias iid, contrastar  $H_0 : F = G$ .
- Independencia:** Dada  $(X_1, Y_1), \dots, (X_n, Y_n) \equiv F$  de vectores bidimensionales aleatorios iid, contrastar  $H_0 : X$  e  $Y$  son independientes.

## Contraste $\chi^2$ de bondad de ajuste (hipótesis nula simple)

Sean  $X_1, \dots, X_n \equiv F$  i.i.d. y  $F_0$  una distribución totalmente especificada. Queremos contrastar  $H_0 : F = F_0$ .

**Ejemplo:** Tiramos un dado 100 veces y obtenemos los resultados siguientes:

$x_i$	1	2	3	4	5	6
Frecuencia	10	20	20	10	15	25

¿Está el dado trucado?

Corresponde a tomar  $F_0$  como la distribución uniforme sobre  $\{1, 2, 3, 4, 5, 6\}$ .

## Contraste $\chi^2$ de bondad de ajuste (hipótesis nula simple)

1. Se definen  $k$  clases  $A_1, \dots, A_k$ . (En el caso del dado, los valores de cada cara.)
2. Se calculan las frecuencias observadas de datos en cada clase

$$O_j = \#\{i : X_i \in A_j\}$$

3. Se calculan las frecuencias esperadas si  $H_0$  fuera cierta  $E_j = np_j$ , donde  $p_j = \mathbb{P}_{F_0}(A_j)$ .
4. Se compara lo observado y lo esperado mediante el **estadístico de Pearson**,

$$T = \sum_{j=1}^k \frac{(O_j - E_j)^2}{E_j} = \sum_{j=1}^k \frac{O_j^2}{E_j} - n.$$

5. Dado un nivel de significación  $\alpha$ , se rechaza  $H_0$  en la región crítica  $R = \{T > c\}$ , donde  $c$  es tal que  $\alpha = \mathbb{P}_{F_0}(T > c)$ .

## Distribución de $T$ bajo $H_0$

**Teorema:** Bajo  $H_0 : F = F_0$ , se verifica

$$\sum_{j=1}^k \frac{(O_j - E_j)^2}{E_j} \rightarrow_d \chi_{k-1}^2.$$

**Justificación intuitiva:**

Si  $n$  grande,  $p_j$  pequeña y  $np_j$  moderado:

$$O_j \equiv \text{Binom}(n, p_j) \cong \mathcal{P}(np_j) \cong \text{N}(\mu = E_j, \sigma^2 = E_j).$$

Además  $\sum_{j=1}^k O_j = n$ , por lo que se pierde un grado de libertad.

# Distribución de $T$ bajo $H_0$

## Demostración:

- ▶ Para  $i = 1, \dots, n$  se define el vector aleatorio  $\xi_i$  con valores en  $\mathbb{R}^k$  que vale  $e_j$  si  $X_i \in A_j$  ( $e_j$  es un vector de ceros salvo un uno en la posición  $j$ ).
- ▶ Se cumple  $\mathbb{E}(\xi_i) = p$  y  $\Sigma \equiv \text{Var}(\xi_i) = P - pp'$ , donde  $p = (p_1, \dots, p_k)'$  y  $P = \text{diag}(p_1, \dots, p_k)$ .
- ▶ Relación entre  $\xi_i$  y estadístico de Pearson:

$$\begin{aligned} \frac{1}{\sqrt{n}} P^{-1/2} \left( \sum_{i=1}^n \xi_i - np \right) &= \left( \frac{(O_1 - E_1)}{\sqrt{E_1}}, \dots, \frac{(O_k - E_k)}{\sqrt{E_k}} \right)' \\ &= P^{-1/2} \sqrt{n} (\bar{\xi} - p). \end{aligned}$$

## Distribución de $T$ bajo $H_0$

- ▶ Por lo tanto

$$T = \|P^{-1/2}\sqrt{n}(\bar{\xi} - p)\|^2$$

- ▶ Por TCL,  $\sqrt{n}(\bar{\xi} - p) \rightarrow_d N_k(0, \Sigma)$  y por el teorema de la aplicación continua,  $T \rightarrow_d \|Y\|^2$ , donde  $Y \equiv N_k(0, P^{-1/2}\Sigma P^{-1/2})$ .
- ▶ La matriz  $V \equiv P^{-1/2}\Sigma P^{-1/2} = \mathbb{I}_k - \sqrt{p}\sqrt{p}'$  es simétrica e idempotente (ejercicio).
- ▶ Por lo tanto  $\|Y\|^2 \equiv \chi_{\text{traza}(V)}^2$  (último ejercicio de la relación de problemas 1), pero  $\text{traza}(V) = k - 1$ .



## Contraste $\chi^2$ de bondad de ajuste (hipótesis nula simple)

Por el resultado anterior, la región crítica del contraste para un nivel de significación  $\alpha$  es:

$$R = \{T > \chi_{k-1; \alpha}^2\}.$$

**Ejemplo:** Tiramos un dado 100 veces y obtenemos los resultados siguientes:

$x_i$	1	2	3	4	5	6
Frecuencia	10	20	20	10	15	25

Contrasta la hipótesis nula de que el dado no está trucado a nivel  $\alpha = 0.05$ . ¿Cuál es el p-valor del contraste?

## Contraste $\chi^2$ de bondad de ajuste ( $H_0$ compuesta)

Sean  $X_1, \dots, X_n \equiv F$  vauid. Ahora  $H_0 : F \in \{F_\theta : \theta \in \Theta \subset \mathbb{R}^r\}$ .

1. Se definen las clases  $A_1, \dots, A_k$  y se calculan las frecuencias observadas  $O_1, \dots, O_k$ .
2. Estimamos  $\theta$  por el método de máxima verosimilitud a partir de  $O_1, \dots, O_k$ . Sea  $\hat{\theta}$  el e.m.v.
3. Se calculan las frecuencias esperadas basadas en el valor del e.m.v., es decir,  $\hat{E}_j = n\hat{p}_j$  donde  $\hat{p}_j = \mathbb{P}_{\hat{\theta}}(A_j)$ .
4. El estadístico de Pearson es

$$T = \sum_{j=1}^k \frac{(O_j - \hat{E}_j)^2}{\hat{E}_j}$$

5. ¿Qué distribución tiene este estadístico bajo  $H_0$ ?

## Contraste $\chi^2$ de bondad de ajuste ( $H_0$ compuesta)

Al estimar  $r$  parámetros se introducen  $r$  nuevas restricciones sobre el vector  $(O_1, \dots, O_r)$ .

Bajo condiciones de regularidad,

$$\sum_{j=1}^k \frac{(O_j - \hat{E}_j)^2}{\hat{E}_j} \rightarrow_d \chi_{k-r-1}^2.$$

### Condiciones:

- ▶  $\Theta$  es un conjunto abierto de un subespacio afín  $r$  dimensional, con  $r < k - 1$ .
- ▶ Si  $\theta_0$  es el verdadero valor de  $\theta$  y  $\hat{\theta}$  es el EMV de  $\theta$ , entonces  $\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d N(0, I(\theta_0)^{-1})$ , donde  $I(\theta_0)$  es la matriz de información de Fisher.
- ▶ Las funciones  $\theta \rightarrow P_{\theta}(A_j)$  son inyectivas y derivables dos veces respecto a  $\theta$ .

La región crítica del contraste para un nivel de significación  $\alpha$  es:

$$R = \{T > \chi_{k-r-1; \alpha}^2\}.$$

## Ejemplo: los bombardeos de Londres

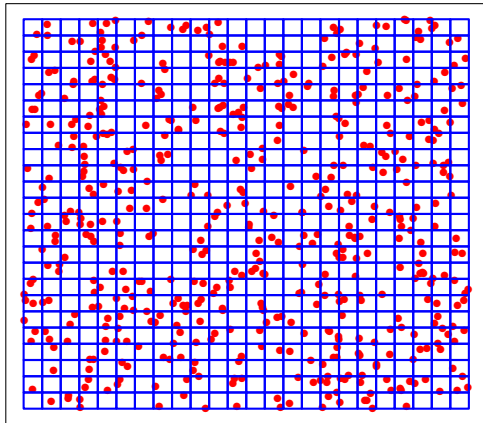


- ▶ Desde el 8 de septiembre de 1944 al 27 de marzo de 1945 fueron lanzados contra Inglaterra unos 1.400 misiles V2.
- ▶ ¿Eran los lugares de impacto en Londres aleatorios?

## Ejemplo: los bombardeos de Londres

- ▶ Clarke, en 1946, dividió un área de  $144 \text{ km}^2$  del sur de Londres en 576 cuadrados de  $0.25 \text{ km}^2$  cada uno.
- ▶ La zona había registrado 537 impactos, por lo que la media era de  $537/576 \approx 0.9323$  misiles por cuadrado.
- ▶ ¿Se ajusta la fórmula de Poisson a la distribución de impactos?

## Ejemplo: los bombardeos de Londres



## Ejemplo: los bombardeos de Londres

The actual results were as follows:

No. of flying bombs per square	Expected no. of squares (Poisson)	Actual no. of squares
0	226.74	229
1	211.39	211
2	98.54	93
3	30.62	35
4	7.14	7
5 and over	1.57	1
	576.00	576

## Test basado en la distribución $\chi^2$

Hipótesis nula: La distribución es Poisson de parámetro  $\lambda$ .

Bajo  $H_0$ , el estimador de máxima verosimilitud de  $\lambda$  es:

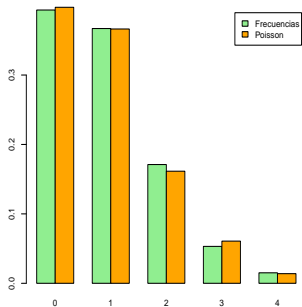
$$\hat{\lambda} = \frac{0 \times 229 + 1 \times 211 + \cdots + 1 \times 7}{576} \approx 0.9323$$

Frecuencias esperadas estimadas bajo  $H_0$ :

$$\hat{E}_k = n\hat{p}_k = 576 \times e^{-\hat{\lambda}} \frac{\hat{\lambda}^k}{k!}$$



$k$	$O_k$	$\hat{p}_k$	$\hat{E}_k$
0	229	0.39	226.74
1	211	0.37	211.34
2	93	0.17	98.54
3	35	0.05	30.62
$\geq 4$	8	0.02	8.71



Estadístico  $\chi^2$  de Pearson:

$$T = \sum_k \frac{(O_k - \hat{E}_k)^2}{\hat{E}_k} = 1.0176$$

Bajo  $H_0$ ,  $T \equiv \chi_3^2$  (5 clases - 1 parámetro estimado -1).

Si  $Y \cong \chi_3^2$ ,  $P\{Y > 1.0176\} \approx 0.797$ .

**El p-valor del contraste es aproximadamente 0.797.**

Al nivel habitual  $\alpha = 0.05$  no se puede rechazar que los datos procedan de una distribución de Poisson.

## Test $\chi^2$ de bondad de ajuste con R

El comando de R que implementa el test  $\chi^2$  es  
`chisq.test(datos,p=...)`

- ▶ `datos`: Es la muestra de que disponemos
- ▶ `p`: Es el vector de probabilidades esperadas.
- ▶ Por defecto, se contrasta la hipótesis de que los datos siguen una distribución uniforme.
- ▶ Se supone que bajo  $H_0$  la distribución está completamente especificada ( $k - 1$  grados de libertad)

Entre otras cosas la función calcula el valor del estadístico y el p-valor del contraste.

Para obtener el estadístico de Pearson:

```
chisq.test(datos,p=...)$statistic
```

## Código R para el ejemplo de los bombardeos

```
res = c(seq(0,4),7)
obs = c(229,211,93,35,7,1)
n = sum(obs)
lambda = sum(res*obs)/n
prob = dpois(res,lambda)
esp = n*prob
# Se agrupan las dos ultimas clases:
obs = c(obs[1:4],sum(obs[5:6]))
prob = c(p[1:4],1-sum(p[1:4]))
esp = c(esp[1:4],n-sum(esp[1:4]))
# Codigo para el grafico de barras:
matriz = rbind(p,obs/n)
rownames(matriz) = c('Frecuencias','Poisson')
barplot(matriz,beside=TRUE,names.arg=c(0:4),legend.text=TRUE,
        col=c('lightgreen','orange'))
# Test chi 2
t = chisq.test(obs,p=prob)$statistic
pvalor = 1 - pchisq(t,3)
```

## Contraste de Kolmogorov-Smirnov

Si  $X_1, \dots, X_n \equiv F$  i.i.d.  $H_0 : F = F_0$ , con  $F_0$  totalmente especificada **y continua**.

Se define la **función de distribución empírica** como

$$F_n(x) = \frac{1}{n} \#\{i : X_i \leq x\}$$

¿A qué distribución de probabilidad corresponde la función de distribución empírica de  $x_1 = 2$ ,  $x_2 = 4$  y  $x_3 = 6$ ?

¿Que distribución tiene la v.a.  $nF_n(x)$ ? ¿A dónde converge  $F_n(x)$ ?

**Teorema (Glivenko-Cantelli):**

$$D_n = \|F_n - F\|_\infty = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \rightarrow_{c.s.} 0.$$

# Contraste de Kolmogorov-Smirnov

La idea es rechazar  $H_0$  en  $R = \{D_n > c\}$ , para un valor crítico  $c$  apropiado.

**Teorema:** La distribución bajo  $H_0$  de  $D_n$  es la misma para cualquier distribución continua  $F_0$ . (Importante: el valor de  $c$  es el mismo para cualquier distribución  $F_0$  continua.)

**Demostración:** Se basa en dos observaciones:

1. **Lema:** Si una v.a.  $X$  tiene distribución continua  $F_0$ , entonces  $F_0(X)$  tiene distribución uniforme en  $(0, 1)$ .

2.  $D_n$  depende de los  $X_i$  solo a través de  $F_0(X_{(i)})$ :

$$D_n = \max \left\{ 0, \max_{i=1, \dots, n} \left( \frac{i}{n} - F_0(x_{(i)}) \right), \max_{i=1, \dots, n} \left( F_0(x_{(i)}) - \frac{i-1}{n} \right) \right\}$$

La distribución de  $D_n$  es fácil de simular y está tabulada.

# Tablas

**Tabla 8** Contraste de Kolmogorov-Smirnov

Valores críticos de  $D = |F_n(x) - F(x)|$  donde  $F_n(x)$  es la distribución muestral de tamaño  $n$  y  $F(x)$  la distribución teórica.

Tamaño muestral $n$	Nivel de significación				
	0,20	0,15	0,10	0,05	0,01
1	0,900	0,925	0,950	0,975	0,995
2	0,684	0,726	0,776	0,842	0,929
3	0,565	0,597	0,642	0,708	0,828
4	0,494	0,525	0,564	0,624	0,733
5	0,446	0,474	0,510	0,565	0,669
6	0,410	0,436	0,470	0,521	0,618
7	0,381	0,405	0,438	0,486	0,577
8	0,358	0,381	0,411	0,457	0,543
9	0,339	0,360	0,388	0,432	0,514
10	0,322	0,342	0,368	0,410	0,490
11	0,307	0,326	0,352	0,391	0,468
12	0,295	0,313	0,338	0,375	0,450
13	0,284	0,302	0,325	0,361	0,433
14	0,274	0,292	0,314	0,349	0,418
15	0,266	0,283	0,304	0,338	0,404
16	0,258	0,274	0,295	0,328	0,392
17	0,250	0,266	0,286	0,318	0,381
18	0,244	0,259	0,278	0,309	0,371
19	0,237	0,252	0,272	0,301	0,363
20	0,231	0,246	0,264	0,294	0,356
25	0,21	0,22	0,24	0,27	0,32
30	0,19	0,20	0,22	0,24	0,29
35	0,18	0,19	0,21	0,23	0,27
>35	$\frac{1,07}{\sqrt{n}}$	$\frac{1,14}{\sqrt{n}}$	$\frac{1,22}{\sqrt{n}}$	$\frac{1,36}{\sqrt{n}}$	$\frac{1,65}{\sqrt{n}}$

$n$  es el tamaño de la muestra.

## Ejemplo

Contrasta a nivel  $\alpha = 0.01$  si la muestra 16, 8, 10, 12, 6 procede de una exponencial de media  $\lambda = 11.5$ .  $F_0(x) = 1 - e^{-\frac{x}{11.5}}$ .

Si  $D_n^+ = \{\frac{i}{n} - F_0(X_{(i)})\}$  y  $D_n^- = \{F_0(X_{(i)}) - \frac{i-1}{n}\}$

$X_{(i)}$	$\frac{i}{n}$	$F_0(X_{(i)})$	$D_n^+$	$D_n^-$
6	0.2	0.41	-0.21	0.41
8	0.4	0.5	-0.1	0.3
10	0.6	0.58	0.02	0.18
12	0.8	0.65	0.15	0.05
16	1	0.75	0.25	-0.05

$D_n = \max\{0, D_n^+, D_n^-\} = 0.41$  y  $D_{6;0.01} = 0.669$ .

No podemos rechazar  $H_0$ .



## Contraste de Kolmogorov-Smirnov con R

Comando de R para el test de Kolmogorov-Smirnov:

```
ks.test(datos,distribucion,parametros)
```

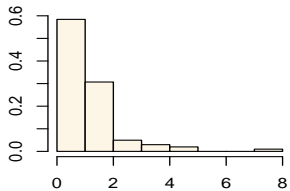
- ▶ `datos`: La muestra de que disponemos
- ▶ `distribucion`: Distribución bajo  $H_0$  (por ejemplo, `pnorm`)
- ▶ `parametros`: Parámetros de la distribución bajo  $H_0$ .

## Datos kevlar

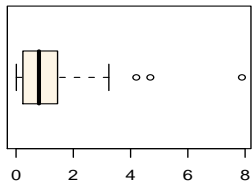
Los datos del fichero `kevlar.RData` corresponden al tiempo hasta el fallo (en horas) de 101 barras de un material utilizado en los transbordadores espaciales, llamado Kevlar49/epoxy, sometidas a un nivel de esfuerzo del 90%.

Los datos han sido tomados de Barlow et al. (1984).

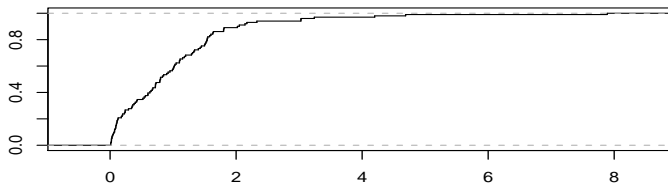
**Histograma**



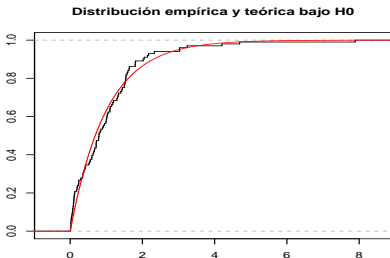
**Diagrama de cajas**



**Distribución empírica**



Vamos a contrastar la hipótesis nula de que los datos tienen distribución exponencial de parámetro  $\lambda = 1$ .



```
ks.test(kev, pexp)
```

```
One-sample Kolmogorov-Smirnov test
```

```
data: kev
```

```
D = 0.087, p-value = 0.4286
```

```
alternative hypothesis: two.sided
```

## Gráficos de probabilidad

- ▶ Es un procedimiento gráfico para ver si una distribución es adecuada para unos datos.
- ▶ Si  $X \equiv F$ , entonces se verifica

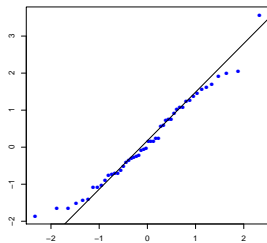
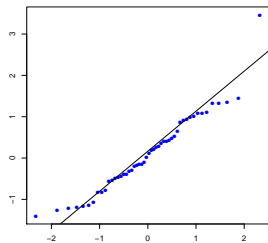
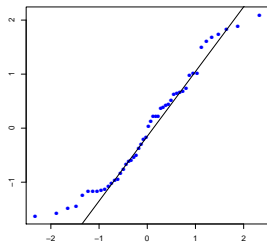
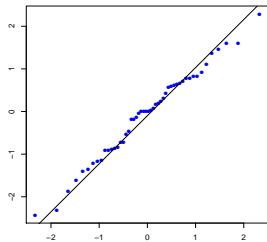
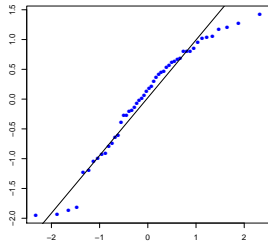
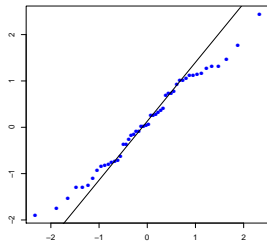
$$\mathbb{E}(F(X_{(i)})) = \frac{i}{n+1}.$$

- ▶ Como consecuencia,  $X_{(i)} \approx F^{-1}(\frac{i}{n+1})$ .
- ▶ Si  $F \equiv N(\mu, \sigma^2)$  y  $\Phi \equiv N(0, 1)$ , entonces

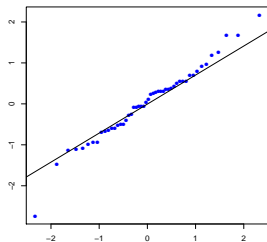
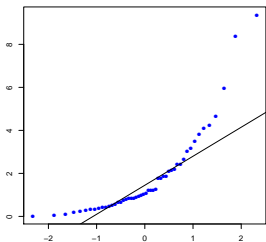
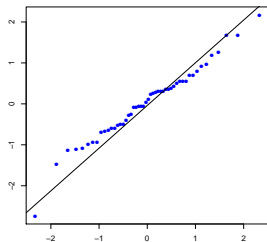
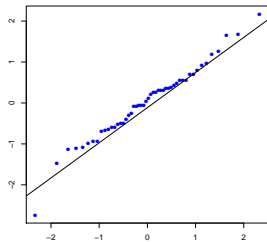
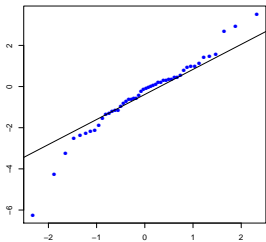
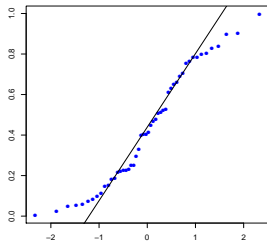
$$X_{(i)} \approx \sigma \Phi^{-1}\left(\frac{i}{n+1}\right) + \mu.$$

- ▶ Se representan los puntos  $(X_{(i)}, \Phi^{-1}(\frac{i}{n+1}))$ .

# Gráficos de probabilidad para 6 muestras normales ( $n = 50$ )



¿Cuáles de estas muestras ( $n = 50$ ) proceden de una distribución normal?



# Contraste de homogeneidad

Para  $p$  muestras independientes  $M_1, \dots, M_p$ ,

$$\left. \begin{array}{l} M_1 \equiv X_{11}, \dots, X_{1n_1} \quad \overset{iid}{\sim} F_1 \\ \vdots \\ M_p \equiv X_{p1}, \dots, X_{pn_p} \quad \overset{iid}{\sim} F_p \end{array} \right\} H_0 : F_1 = F_2 = \dots = F_p$$

- ▶ Dividimos los datos en clases  $A_i$  y consideramos las frecuencias observadas  $O_{ij} = \#\{\text{datos de } M_j \text{ en } A_i\}$
- ▶ Bajo  $H_0$ ,  $O_{ij} \equiv B(n_j, p_i)$ , con  $p_i = \mathbb{P}_{H_0}(A_i)$ .
- ▶ Por lo tanto,  $E_{ij} = n_j p_i$ .

## Tabla de contingencia

Las dos matrices siguientes se comparan con el estadístico de Pearson:

	$M_1$	$\dots$	$M_p$		$M_1$	$\dots$	$M_p$
$A_1$	$O_{11}$	$\dots$	$O_{1p}$	$A_1$	$E_{11}$	$\dots$	$E_{1p}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$A_k$	$O_{k1}$	$\dots$	$O_{kp}$	$A_k$	$E_{k1}$	$\dots$	$E_{kp}$



## Contraste de homogeneidad

- ▶ Bajo  $H_0$  se verifica

$$\sum_j \sum_i \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \rightarrow_d \chi_{p(k-1)}^2.$$

- ▶ Dado que no conocemos las probabilidades  $p_i$ , tenemos que estimarlas. Bajo homogeneidad,

$$\hat{p}_i = \frac{\sum_j O_{ij}}{\sum_j n_j} \implies \hat{E}_{ij} = n_j \hat{p}_i = \frac{O_{i \cdot} \cdot O_{\cdot j}}{n}.$$

- ▶ El estadístico en la práctica (y su dist. bajo  $H_0$ ) es:

$$T = \sum_i \sum_j \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} \rightarrow_d \chi_{(p-1)(k-1)}^2$$

- ▶ La región crítica es

$$R = \{T > \chi_{(k-1)(p-1), \alpha}^2\}$$

## Jane Austen

Al morir la escritora Jane Austen (1775–1817) dejó sin concluir su novela *Sanditon*. Desde entonces, varios autores han completado de diferentes formas la novela.

Los datos siguientes (Rice, p. 488) corresponden a la frecuencia de uso de algunas palabras en una muestra de novelas de Austen y en la continuación escrita por un imitador.

Palabra	a	an	this	that	with	without
Imitador	83	29	15	22	43	4
Austen	434	62	86	236	161	38

Sobre la base de estos datos, ¿podemos distinguir entre el estilo de Austen y el estilo del imitador?

## Solución

Frecuencias esperadas:

Palabra	a	an	this	that	with	without
Imitador	83.54	14.7	16.32	41.69	32.96	6.79
Austen	433.46	76.3	84.68	216.31	171.04	35.21

Diferencias estandarizadas al cuadrado:

Palabra	a	an	this	that	with	without
Imitador	0	13.90	0.11	9.30	3.06	1.14
Austen	0	2.68	0.02	1.79	0.59	0.22

Estadístico de Pearson:  $T = 0 + 13.90 + \dots + 0.22 \approx 32.81$

Bajo  $H_0$   $T$  sigue aproximadamente una distribución  $\chi^2$  con 5 gl.

Mirando las tablas de la  $\chi^2$ , vemos que  $\chi_{5,0.05}^2 = 11.07$

¿Cuál es la conclusión?

## Solución con R

```
> austen = matrix(c(83,434,29,62,15,86,22,  
> + 236,43,161,4,38),2)  
> austen  
      [,1] [,2] [,3] [,4] [,5] [,6]  
[1,]   83   29   15   22   43    4  
[2,]  434   62   86  236  161   38  
> chisq.test(austen)
```

Pearson's Chi-squared test

data: austen

X-squared = 32.8096, df = 5, p-value = 4.106e-06

# Contraste de independencia

Sean  $(X_1, Y_1), \dots, (X_n, Y_n) \stackrel{iid}{\sim} F$ . Queremos contrastar  $H_0 : X, Y$  son v.a. independientes.

Los datos se suelen dar en forma de tabla de contingencia (agrupados):

	$B_1$	$\dots$	$B_p$
$A_1$	$O_{11}$	$\dots$	$O_{1p}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
$A_k$	$O_{k1}$	$\dots$	$O_{kp}$

Frecuencias esperadas y su estimación:

$$E_{ij} = np_{ij} = nP(X \in A_i, Y \in B_j) \stackrel{H_0}{=} nP(X \in A_i)(Y \in B_j)$$
$$\hat{E}_{ij} = n \frac{O_{i\cdot}}{n} \frac{O_{\cdot j}}{n} = \frac{O_{i\cdot} O_{\cdot j}}{n}.$$

Formalmente, el contraste es exactamente igual al contraste de homogeneidad.