Dpto. de Economía Cuantitativa
Universidad Complutense de Madrid

Introducción a la Econometría

Tema 7 — Repaso a la contrastación de Hipótesis Estadísticas.

Métodos No Paramétricos

Marcos Bujosa y Gustavo A. Marrero

Material de apoyo para el curso *Introducción a la Econometría* de la licenciatura en Economía de la Universidad Complutense de Madrid.

Actualizado el: 22 de mayo de 2007 Versión 4.1

Índice

Este es un material de apoyo a las clases. En ningún caso sustituye a los libros de texto que figuran en el programa de la asignatura; textos que el alumno debe estudiar para afrontar el examen final con ciertas garantías de éxito.

El programa se cubre con los siguientes capítulos de libro de texto Novales (1997)[1]:

[1] Otros excelentes manuales en castellano son Peña (2001), Peña (2002) y Peña y Romo (1997).

Capítulos 1 a 3: Estos temas han sido cubiertos en asignaturas anteriores, y debido a su bajo nivel de complejidad no se verán en clase (aunque forman parte del programa).

Capítulos 4 a 6: Estos temas han sido cubiertos en las asignaturas Estadística I y II. Se realizará un breve repaso en clase (una semana o semana y media como máximo), asumiendo que el alumno es capaz de preparar por su cuenta esta parte.

Capítulos 7 y 8: completos

Capítulo 9: secciones 9.4 a 9.6

Capítulos 10 y 12: completos

1. Introducción a los contrastes de hipótesis no paramétricos

⇑ **Contrastes de hipótesis: paramétricos** 1

A.- Suponemos mismo modelo de probabilidad para H_0 y H_1:

$$\Phi_0 = \{ f_X(x; \boldsymbol{\theta}), \ \boldsymbol{\theta} \in \boldsymbol{\Theta_0}, \ x \in \mathbb{R}_X \}$$

$$\Phi_1 = \{ f_X(x; \boldsymbol{\theta}), \ \boldsymbol{\theta} \in \boldsymbol{\Theta_1}, \ x \in \mathbb{R}_X \}$$

$\boldsymbol{\Theta_0} \cup \boldsymbol{\Theta_1} = \boldsymbol{\Theta}$

B.- Suponemos Muestra Aleatoria Simple (indep. e idéntica distrib.)

C.- Elegimos estadístico y dividimos espacio muestral (el soporte conjunto) en dos subconjuntos

⇑ **Contrastes de hipótesis: No paramétricos** 2

Tipos

1. Distribución

$$H_0 : \boldsymbol{X} \sim f_{\boldsymbol{X}}(\boldsymbol{x})$$
$$H_1 : \boldsymbol{X} \nsim f_{\boldsymbol{X}}(\boldsymbol{x})$$

2. Homogeneidad entre muestras

$$H_0 : X_k \ \text{ son ID}$$
$$H_1 : X_k \ \text{ no son ID}$$

para $k = 1, \ldots, n$

3. Independencia

$$H_0 : X_k \ \text{ son I}$$
$$H_1 : X_k \ \text{ no son I}$$

para $k = 1, \ldots, n$

2. Contrastes de ajuste a una Distribución

⇑ **Ajuste a una distribución: Hipótesis H_0 y H_1** 3

Suponiendo $\boldsymbol{X} = \{X_1, \ldots, X_n\} \sim$ I.I.D., y por tanto

$$\boldsymbol{X} \sim f_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{n} f_X(x) = (f_X(x))^n ;$$

se contrasta:

$$H_0 : \ \boldsymbol{X} \sim f_{\boldsymbol{X}}(\boldsymbol{x})$$

frente a

$$H_1 : \ \boldsymbol{X} \nsim f_{\boldsymbol{X}}(\boldsymbol{x})$$

Contraste de Normalidad: Jarque-Bera 4

$$H_0: \ \boldsymbol{X} \sim \text{Normal} \quad \longleftrightarrow \quad H_1: \ \boldsymbol{X} \nsim \text{Normal}$$

suponiendo $\boldsymbol{X} = \{X_1, \ldots, X_n\} \sim$ I.I.D.

$$JB = n\left[\frac{AS^2}{6} + \frac{(K-3)^2}{24}\right] \underset{H_0}{\sim} \chi^2_{(2)}$$

Rechazamos si $JB(\mathbf{x}) > \chi^2_{\alpha,2}$

A resolver en clase

Test. Conteste a las siguientes cuestiones.

1. Queremos contrastar si los rendimientos de un determinado activo financiero tienen distribución normal. Para ello tomamos una muestra de tamaño 69. Para esta muestra calculamos el coeficiente de asimetría y el de curtosis y resultan ser 0.2 y 4.2, respectivamente. Para contrastar empleamos el estadístico Jarque-Bera. ¿Cuál de las siguientes afirmaciones es correcta?

(a) El p-valor del contraste es 0.10, por lo que a un nivel de significación del 0.05 rechazamos que es normal.

(b) El p-valor del contraste es 0.10, por lo que a un nivel de significación del 0.05 **no** rechazamos que es normal.

(c) El p-valor del contraste es 5.99.

(d) El p-valor del contraste es 0.05.

Ajuste a una distribución: Kolmogorov-Smirnov 5

Sólo válido para distribuciones continuas.

1. Ordenar los valores muestrales
$$x_{(1)} \leq x_{(2)} \ldots \leq x_{(n)}$$
2. Calcular la función de distribución empírica $F_n(x)$
$$F_n(x) = \frac{1}{n}\sum_{i=1} I_{\{x_i \leq x\}}$$
3. Calcular la discrepancia máxima
$$D_n = \sup |F_n(x) - F_X(x)|$$
cuya distribución está tabulada.
4. Fijado α, si D_n excede el valor de las tablas rechazamos H_0

La *función indicador* se define del siguiente modo:

$$I_{\{\text{condición}\}} = \begin{cases} 1 & \text{si la condición es cierta} \\ 0 & \text{si la condición es falsa} \end{cases}$$

Ejemplo 1. [**Contraste Kolmogorov-Smirnov (ajuste a una distribución exponencial):**]
$\mathbf{x} = \{16, 8, 10, 12, 6, 10, 20, 7, 2, 24\}$, $\overline{x} = 11.5 \quad (\lambda = \frac{1}{11.5})$, $\ n = 10$;

H_0: $X \sim \exp(11.5)$; por tanto $H_0: \ F_X(x) = 1 - e^{-\frac{x}{11.5}}$.

x	$F_n(x)$	$F_X(x)$	$D_n(x)$
2	0.1	0.16	0.06
6	0.2	0.41	0.21
7	0.3	0.46	0.16
8	0.4	0.50	0.10
10	**0.6**	0.58	0.02
12	0.7	0.65	0.05
16	0.8	0.75	0.05
20	0.9	0.82	0.08
24	1.0	0.88	0.12

Entonces $D_n = 0.21$. En las tablas $D_{0.2,10} = 0.322$ y $D_{0.1,10} = 0.368$

Añadir gráfico;

```
x=[2 6 7 8 10 12 16 20 24]';
Fempirica=[.1 .2 .3 .4 .6 .7 .8 .9 1]'; Fteorica=[1-exp(-(x/11.5))];
hold on, stairs(x,[Fempirica]);plot(x,[Fteorica],'-*');
plot(x, Fteorica-0.3222,'g');plot(x, Fteorica+0.3222,'g'); hold off
```

Ejercicios de muestra Novales (1997, Ejemplo 12.4 Pag. 439)

Comentario.

- Si fuera necesario estimar uno o más parámetros de $\boldsymbol{\theta}$ la distribución de D_n sería sólo aproximada.
- Es conservador (tiende a no rechazar H_0)
- Permite construir bandas de confianza. Si H_0: $F_X(x)$ es correcta, con confianza $(1-\alpha)$

$$F_X(x) \in \left[F_n(x) \pm D_{\alpha/2,n}.\right]$$

⇑ **Ajuste a una distribución: Chi-cuadrado** 6

Es aplicable a distribuciones tanto discretas como continuas

Suponemos $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ *m.a.s.*, donde $n \geq 25$.

1. agrupamos datos en $k \geq 5$ intervalos o subconjuntos I_i; $i = 1, \ldots, k$ (partición de $\mathbb{R}_X$)[a]; llamamos O_i a la frecuencia absoluta en cada I_i.
2. $p_i = P_{H_0}(X \in I_i)$; por tanto, $\boxed{\sum_{i=1}^k p_i = 1}$; y llamamos $e_i = np_i$ a la frecuencia esperada asociada al intervalo i bajo H_0. [b]
3. El estadístico es
$$g(\boldsymbol{X}) = \sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i} \overset{a}{\sim} \chi^2.$$
con $k-1$ grados de libertad.

[a] con aprox. mismo nº de datos en cada intervalo

[b] procuraremos tener nº intervalos y de tamaño tal que para cada subconjunto $np_i > 5$

Nota 1. El concepto de partición de un conjunto tiene un significado preciso: $\cup_{i=1}^k I_i = \mathbb{R}_X$ y $I_i \cap I_j = \emptyset$.

⇑ **Ajuste a una distribución: Chi-cuadrado** 7

Si el vector $\boldsymbol{\theta}$ de $f_X(x;\boldsymbol{\theta})$ es conocido (H_0 simple)

$$g(\boldsymbol{X}) \overset{a}{\sim} \chi^2_{k-1}$$

(una restricción lineal: $\sum O_i = n$;)

Si r parámetros de $\boldsymbol{\theta}$ son estimados,

$$g(\boldsymbol{X}) \overset{a}{\sim} \chi^2_{k-r-1}$$

(cada parámetro estimado supone una restricción adicional).

Rechazamos H_0 si

$$g(\mathbf{x}) = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} \geq \chi^2_{\alpha,k-r-1}$$

Justificación de la distribución del contraste Chi-cuadrado Sea I_i un intervalo; entonces $O_i \sim$ Binomial $(n,\ p_i)$ y por tanto $\mathrm{E}(O_i) = np_i = e_i$; y $\mathrm{Dt}(O_i) = \sqrt{np_i(1-p_i)}$.

Si n grande y p pequeño, O_i será aproximadamente una Poisson (λ) con $\lambda = np_i$ (véase Sección B del apéndice).

Puede considerarse una Poisson (λ) como suma de λ variables Poisson (1) I.I.D. [2] (recuérdese que $\mathrm{E}(\mathrm{Poisson}(\lambda)) = \lambda = \mathrm{Var}(\mathrm{Poisson}(\lambda))$).

Si λ es suficientemente grande (en la práctica $\lambda > 5$) podemos utilizar la aproximación normal mediante el Teorema Central del Limite. Entonces:

$$\frac{O_i - np_i}{\sqrt{np_i}} \overset{a}{\sim} N(0,1)$$

puesto que $\sum O_i = n$ hay restricción lineal (1 grado de libertad menos)

$$\sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i} = \sum_{i=1}^{k} \frac{(O_i - np_i)^2}{n_i p_i} \overset{a}{\sim} \chi^2_{k-1}$$

Ejercicios de muestra Novales (1997, Ejemplo 12.1 y 12.2 Pag. 436)

A resolver en clase

Ejercicio 1. De 1000 encuestados, 680 son favorables a que bajen los impuestos sobre la gasolina y 320 no lo son. Un político manifiesta que el 75 % del total de la población es favorable a la propuesta y un 25 % no lo es. Contraste la veracidad de la afirmación del político según la encuesta realizada [Pista: lleve a cabo un contraste chi-cuadrado de ajuste a una distribución teórica].

A resolver en clase

Ejercicio 2. (Consta de 4 apartados)

La variable aleatoria X mide los rendimientos semanales de un determinado Fondo de Inversión. La información que presentamos a continuación resume las rentabilidades de los últimos 100 días. Los datos de las rentabilidades han sido previamente estandarizadas, por lo que tienen media cero y varianza uno. Sus realizaciones han sido ordenadas en 7 intervalos (pérdidas mayores al 1.5 %, pérdidas de entre el uno y el uno y medio, etc.).

Rentabilidades	< -1.5	$[-1.5; -1)$	$[-1; -0,5)$	$[-0,5; 0,5]$	$(0,5; 1]$	$(1; 1.5]$	> 1.5
Nº de días	9	6	12	41	19	8	5

(a) Queremos contrastar (al 10 % de significación) si los rendimientos (estandarizados) siguen una distribución N(0,1). Lleve a cabo el contraste por los tres métodos vistos en el curso (Jarque-Bera, Kolmogorov-Smirnov, y Chi cuadrado) y discuta si cambian las conclusiones de un contraste a otro (**Dato**: el coeficiente de curtosis de los datos estandarizados es 3,23 y el de asimetrıa -0,24).

(b) Suponga que, a partir de los datos originales sin estandarizar, comprueba que la media muestral es 1,15 y su cuasi-varianza muestral es 0,04; y teniendo en cuenta la conclusión a la que ha llegado en el apartado anterior, quiere contrastar que la esperanza poblacional de la variable X es igual a 1 frente a que es mayor que 1. Defina correctamente la hipótesis nula y la alternativa de este contraste y la región crítica más apropiada. Resuelva el contraste para un nivel de significación del 5 % y del 10 %.

(c) Halle el p-valor del contraste anterior. ¿Qué significado tiene el p-valor? ¿Qué implicaciones tiene a la hora de rechazar o no la hipótesis nula?

(d) Explique (sin necesidad de resolver el contraste) cómo cambiaría el planteamiento del contraste del apartado (b) si la conclusión del apartado (a) hubiese sido la contraria a la que usted ha llegado.

A resolver en clase

Ejercicio 3. Contraste si la siguiente muestra proviene de una distribución Uniforme (0, 1), empleando el contrates Kolmogorov-Smirnov con tres intervalos de misma amplitud.

0.15 0.77 0.70 0.77 0.24 0.05 0.26 0.94 0.75 0.45

A resolver en clase

Ejercicio 4. Suponga una muestra de 40 datos con los siguientes estadísticos: media muestral de 1/2, varianza muestral de 3, coeficiente de asimetría de 1/2 y Kurtosis de 2,5. ¿Cual es el p-valor del contraste Jarque-Bera de esta muestra? ¿Qué concluiría a raíz del p-valor obtenido para la muestra?

[2](Novales, 1997, pp 317)

3. Contrastes de Homogeneidad entre muestras

⇑ **Ajuste a una distribución: Hipótesis H_0 y H_1** **8**

Se contrasta:

$$H_0: \; f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{\boldsymbol{Y}}(\boldsymbol{y})$$

frente a

$$H_1: \; f_{\boldsymbol{X}}(\boldsymbol{x}) \neq f_{\boldsymbol{Y}}(\boldsymbol{y})$$

suponiendo

$$\boldsymbol{X}_{(n)} = \{X_1, \dots, X_n\} \sim \text{ I.I.D.}$$
$$\boldsymbol{Y}_{(m)} = \{Y_1, \dots, Y_m\} \sim \text{ I.I.D.}$$

y $\boldsymbol{X}_{(n)}$ independiente de $\boldsymbol{Y}_{(m)}$

⇑ **Contraste de Homogeneidad: Kolmogorov-Smirnov** **9**

Estadístico:

$$D_{n,m} = \sup |F_n(x) - F_m(y)|$$

Cuya distribución está tabulada.
Cuando n y m son grandes

$$\text{Nivel crítico: } D_{\alpha,n,m} \approx k\sqrt{\frac{n+m}{nm}},$$

donde $k = 1.22;\; 1.36;\; 1.63$; para α igual a 10 %, 5 % y 1 % respectivamente.

Kolmogorov-Smirnov: Ejercicios de muestra Novales (1997, Ejemplo 12.10 Pag. 448)

⇑ **Contraste de Homogeneidad: Chi cuadrado** **10**

Es aplicable a distribuciones tanto discretas como continuas.
Datos clasificados por dos carácterísticas (A y B). Estimamos:

1. Probabilidades de cada grupo i-ésimo de A (columnas) suponiendo igual distribución: $p_i = \dfrac{\sum c_{\bullet i}}{m+n}$; por tanto, $\boxed{\sum_i p_i = 1}$.
2. Frecuencias esperadas $e_{ij} = n_j p_i$ de cada caso:
   $e_{ij} = (\text{observaciones fila}_j) \times p_i$; por tanto $\boxed{\sum_j e_{ij} = n_i}$,
   donde $n_i =$ observaciones fila$_i$.
3. El estadístico es
   $$g(\boldsymbol{X}) = \sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i} \overset{a}{\sim} \chi^2$$
   con (n$^{\text{o}}$ de filas-1)×(n$^{\text{o}}$ de columnas-1) grados de libertad.

A resolver en clase

Ejercicio 5. Se pretende contrastar si la excelencia de los alumnos se mantiene entre dos cursos consecutivos (curso 2005 y curso 2006). Para ello se dividen las calificaciones de 1200 alumnos en las siguientes categorías: "1: suspenso", "2: aprobado", "3: aprobado con nota". La información muestral se resume del siguiente modo:

	(1) suspensos	(2) aprobados	(3) con nota
2005	240	190	70
2006	310	270	120

a) Calcule la tabla de frecuencias absolutas esperadas bajo H_0. b) Para esta muestra, el estadístico Chi cuadrado arroja un valor aproximado de 2.7 ¿Cuanto vale el p-valor? (emplee el valor más próximo de los que aparecen en las tablas). c) Lleve a cabo el contraste Chi cuadrado para niveles de significación del 5 % y del 10 %.

A resolver en clase

EJERCICIO 6. Calcule el valor del contraste de homogeneidad de Kolmogorov-Smirnov para dos muestras con los datos del ejercicio anterior.

EJERCICIO 7. **Chi cuadrado:** Contrastar si los grupos sanguíneos se distribuyen de manera igual entre hombres y mujeres. Muestra de 12000 personas; 5000 hombre y 7000 mujeres (Novales, 1997, Ejemplo 12.14 Pag. 454):

	Grupo 0	Grupo A (1)	Grupo B (2)
Mujeres X	1200	3100	2700
Hombres Y	700	2400	1900

A resolver en clase

EJERCICIO 8. Hemos encuestado a 500 empresarios de Madrid y a 450 de Barcelona sobre cuáles son sus expectativas de comercio para los próximos 12 meses con respecto a su situación actual. Las 5 posibles respuestas son: mucho peor (-2), peor (-1), similar (0), mejor (1) y mucho mejor (2). El resultado de la encuesta se resume en la siguiente tabla:

	mucho peor (-2)	peor (-1)	similar (0)	mejor (1)	mucho mejor (2)
Madrid	100	150	125	100	25
Barcelona	68	157	112	68	45

Queremos contrastar si la visión que tienen los empresarios madrileños acerca de su actividad económica futura es diferente de la de los catalanes. Proponga y resuelva dos contrastes apropiados (Kolmogorov y Chi-cuadrado) para abordar esta pregunta. Para cada contraste, use niveles de significación del 1 % y del 10 %. Comente los resultados.

3.1. Contraste de Wilcoxon

Suponga que X_A y X_B representan las ventas semanales de un conjunto de farmacias, las primeras con cartel luminoso y las segundas sin el. Se desea contrastar si la distribución de las ventas es homogenea entre los dos grupos.

Estab.	X_A	X_B	$X_A - X_B$	$\lvert X_A - X_B\rvert$	rango	r. con signo
1	78	78	0	0	—	—
2	24	24	0	0	—	—
3	64	62	+2	2	1	+1
4	45	48	-3	3	2	-2
5	64	68	-4	4	(3)3.5	-3.5
6	52	56	-4	4	(4)3.5	-3.5
7	30	25	+5	5	5	+5
8	50	44	+6	6	6	+6
9	64	56	+8	8	7	+7
10	50	40	+10	10	(8)8.5	+8.5
11	78	68	+10	10	(9)8.5	+8.5
12	22	36	-14	14	10	-10
13	84	68	+16	16	11	+11
14	40	20	+20	20	12	+12
15	90	58	+32	32	13	+13
16	72	32	+40	40	14	+14
		SUMAS	$T^+ = 86$;	$T^- = 19$;		$T = T^+ - T^- = 67$
				Nº de rang.		$N = 14$

Dos colas Al 5 % los valores críticos mínimo y máximo para $N = 14$ son 25 y 80 respectivamente. Para esta muestra $T^- = 19$ y $T^+ = 86$ están fuera de dichos límites, por lo que podemos rechazar H_0.

Además, bajo H_0:

$$\mathrm{E}\,(T) = 0, \quad \mathrm{Var}\,(T) = \frac{n(n+1)(2n+1)}{6};$$

y

$$\mathrm{E}\left(T^+\right) = \frac{n(n+1)}{4}, \quad \mathrm{Var}\left(T^+\right) = \frac{n(n+1)(2n+1)}{24}$$

si la muestra es suficientemente grande (Nº de rangos ≥ 25) es posible emplear la aproximación Normal del estadístico, es decir

$$\frac{T}{\sqrt{\mathrm{Var}\,(T)}} \overset{a}{\sim} \mathrm{N}\,(0\,,\,1)\,; \quad \text{y} \quad \frac{T^+ - \mathrm{E}\,(T^+)}{\sqrt{\mathrm{Var}\,(T^+)}} \overset{a}{\sim} \mathrm{N}\,(0\,,\,1)$$

En este caso los valores críticos serán

$$a = 1.96 * \sqrt{\frac{14*(15)*(29)}{6}} = 62.44$$

y

$$b = -62.44$$

por lo que $T = 67$ de nuevo queda fuera, por lo que rechazamos H_0.

Una sóla cola Al 5 % de significación el valor críticos máximo para $N = 14$ es 74. En este caso T^+ también supera dicho limite por lo que rechazamos H_0.

Si empleamos la aproximación Normal el valor crítico será

$$a = \frac{14*15}{4} + 1.64 * \sqrt{\frac{14*(15)*(29)}{24}} = 82.124$$

$T^+ = 86 > a$ por lo que rechazamos H_0.

3.2. Contraste de Mann-Whitney

Ejercicio 5.7 del libro de problemas (pag 230)

H_0: Misma distribución.

Goles	rangos	
1	$_{(3)}$ 4.5	
2	$_{(7)}$ 8	
0	$_{(1)}$ 1.5	
4	13	
3	$_{(10)}$ 11	
0	$_{(2)}$ 1.5	
1	$_{(4)}$ 4.5	
3	$_{(11)}$ 11	$R_1 = 55$
2	$_{(8)}$ 8	
2	$_{(9)}$ 8	
1	$_{(5)}$ 4.5	
3	$_{(12)}$ 11	
6	12	
1	$_{(6)}$ 4.5	$R_1 = 50$

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1; \qquad U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

$U = \text{mín}(U_1, U_2)$; Si U menor que valor critico, se rechaza H_0.

En este ejercicio

$$U_1 = 8 \cdot 6 + \frac{8(8+1)}{2} - 55 = 29; \qquad U_2 = 8 \cdot 6 + \frac{6(6+1)}{2} - 50 = 19$$

Así que $U = 19$. El valor crítico al 5 % de significación es 9, por lo que no rechazamos H_0.

Existe una aproximación Normal.

$$\mathrm{E}\,(U) = \frac{n_1 n_2}{2}; \qquad \mathrm{Var}\,(U) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

En este ejemplo

$$\frac{8 \cdot 6}{2} = 24, \qquad \frac{8 \cdot 6(8+6+1)}{12} = 60;$$

así pues, aquí podemos suponer que

$$U \overset{a}{\sim} \mathrm{N}\,(24\,,\,60)\,.$$

y el estadístico estandarizado es

$$\frac{19-24}{\sqrt{60}} = -0.64 \quad \in (-1.96;\ 1.96)$$

por lo que tampoco rechazamos H_0 al emplear esta aproximación.

4. Contrastes de Independencia

Contraste de Independencia: Chi cuadrado 11

Datos clasificados por dos características (A y B). Estimamos:

1. Probabilidades de cada grupo i-ésimo de A (columnas) suponiendo igual distribución: $p_i = \dfrac{\sum c_{\bullet i}}{m+n}$; por tanto, $\boxed{\sum p_i = 1}$.
2. Probabilidades de cada grupo j-ésimo de B (filas) suponiendo igual distribución: $p_j = \dfrac{\sum c_{j\bullet}}{m+n}$; por tanto, $\boxed{\sum p_j = 1}$.
3. Frecuencias esperadas $e_{ij} = p_i \cdot p_j \cdot (m+n)$ de cada caso suponiendo indep.: (producto de la probabilidades marginales = probabilidad conjunta)
4. El estadístico es
$$g(\boldsymbol{X}) = \sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i} \overset{a}{\sim} \chi^2$$
con (nº de filas-1)×(nº de columnas-1) grados de libertad.

A resolver en clase

Ejercicio 9. Un canal de televisión realiza una encuesta a 5000 menores de 30 años, y la siguiente tabla de contingencia resume la información acerca del gusto por las películas de Van Damme:

	gusta	no gusta
chicos	348	3152
chicas	82	1418

Calcule el valor del estadístico Chi-cuadrado para contrastar la independencia entre el gusto por las películas y el sexo. ¿Para qué niveles de de significación rechazaría H_0?

Ejercicio 10. El p-valor de un determinado contraste es 0.12. ¿Qué nos dice esta información?

Ejercicio 11. La siguiente tabla de contingencia muestra la información recogida por una encuesta sobre la intención de voto en las próximas elecciones europeas. La variable X toma valor 1 si su intención es votar a un partido de centro-izquierdas y 0 si su intención es votar a un partido de centro-derechas; la variable Y es igual a 1 si es hombre y 0 si es mujer. Queremos contrastar si la intención de voto depende del sexo de los votantes mediante un contraste Chi-cuadrado. ¿Cuáles serían las frecuencias teóricas absolutas para llevar a cabo este contraste?

X/Y	mujer	hombre
centro-derechas	50	75
centro-izquierdas	65	80

A resolver en clase

Ejercicio 12. Queremos contrastar la independencia entre el nivel de ingresos de una familia y su ideología política. Para ello encuestamos a un total de 400 familias y el siguiente cuadro resume la información obtenida:

	derecha	izquierda
$<$ 18 mil euros	80	110
$\geq$18 mil euros	125	85

A partir de esta información muestral, ¿cuáles son las frecuencias teóricas absolutas para llevar a cabo un contraste Chi-cuadrado?

A resolver en clase

EJERCICIO 13. La proporción de alumnos aprobados en junio fue de 150 de los 350 presentados por la mañana y de 75 de los 180 presentados por la tarde. Queremos hacer un contraste no paramétrico chi-cuadrado de independencia entre el turno y aprobar el examen. ¿Cuáles son las frecuencias absolutas teóricas para llevar a cabo este contraste?
(NO resuelva el contraste, limítese a contestar lo que se le pide)

Chi cuadrado: Ejercicios de muestra Novales (1997, Ejemplo 12.18 y 12.19 Pags. 462 a 465)

A resolver en clase

Test. En una ciudad hay dos equipos: 1 y 2, y dos zonas: N(Norte) y S (Sur). Un investigador piensa que ser seguidor del equipo 1 ó del 2 depende de que se viva en la zona N o S. Encuestando a 100 personas, obtiene los resultados que se muestran en la siguiente tabla:

	del equipo 1	del equipo 2	total
zona N	25	30	55
zona S	5	40	45
total	30	70	100

El investigador utiliza un estadístico χ^2 para contrastar H_0 de independencia entre ambas cualidades (equipo del que se es seguidor y zona en la que se vive) contra H_1 de no independencia.

1. A partir de las frecuencias absolutas teóricas bajo H_0, se tiene:
   (a) hay los mismos seguidores del equipo 1 que del 2.
   (b) hay los mismos seguidores del equipo 1 en la zona N que en la S.
   (c) en la zona N hay los mismos seguidores del equipo 1 que del 2.
   (d) ninguna de las anteriores.

2. Con los datos de la tabla anterior se tiene un valor (aproximado) del estadístico de 13.9 y por tanto:
   (a) se rechaza H_0 con una significación del 5 %.
   (b) se rechaza H_0.
   (c) no se rechaza H_0 con una significación del 5 %.
   (d) ninguna de las anteriores.

3. El contraste anterior no podría realizarse si en la población (la ciudad) objeto de estudio:
   (a) no hay seguidores del equipo 1.
   (b) hay los mismos seguidores del equipo 1 que del 2 en la zona N.
   (c) no hay seguidores del equipo 1 de la zona N.
   (d) hubiera podido realizarse en cualquiera de los tres casos anteriores.

Chi cuadrado (tablas de contingencia): Ejercicios de muestra Novales (1997, Ejemplo 12.18 Pag. 462)

4.1. Contraste de correlación por rangos de Spearman

Ejemplo 12.16 del libro de Novales (pag. 459)

Alumno	Macro	Mat	Macro	Mat	d_i	d_i^2
1	70	64	7	9	-2,0	4,0
2	82	90	4	1	3,0	9,0
3	54	36	11	15	-4,0	16,0
4	91	86	2	2,5	-0,5	0,25
5	64	52	10	12,5	-2,5	6,25
6	87	76	3	6	-3,0	9,0
7	32	40	14	14	0,0	0,0
8	74	82	6	5	1,0	1,0
9	66	54	9	11	-2,0	4,0
10	42	58	13	10	3,0	9,0
11	52	66	12	8	4,0	16,0
12	30	52	15	12,5	2,5	6,25
13	78	84	5	4	1,0	1,0
14	92	86	1	2,5	-1,5	2,25
15	68	72	8	7	1,0	1,0
suma			120	120	0	85

$$r_s = 1 - \frac{6\sum_{1=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{6\times 85}{15(15^2-1)} = 0.8482; \qquad r_s\sqrt{\frac{n-2}{1-{r_s}^2}} \sim t_{n-2}$$

En este caso $r_s\sqrt{\frac{n-2}{1-r_s^2}} = 5.774$ que queda fuera de las tablas, por lo que podemos rechazar H_0 (ausencia de relación) para casi cualquier nivel de significación.

5. Problemas y ejercicios

A resolver en clase

Ejercicio 14. Suponga que durante 100 días medimos la rentabilidad de dos fondos de inversión tecnológicos ofrecidos por dos entidades financieras distintas, la entidad A y la entidad B. Se definen intervalos de rentabilidades (menos de $-3\,\%$, entre $-3\,\%$ y menos de $-2\,\%$, . . . , entre dos y tres por ciento, y más de tres por ciento). En la siguente tabla figura el número de días que cada fondo ha arrojado una determinanda rentabilidad.

	<-3	$[-3;-2)$	$[-2;-1)$	$[-1;0)$	$[0;1)$	$[1;2)$	$[2;3)$	>3
A	10	4	12	22	22	12	10	8
B	5	10	15	20	20	16	10	4

(a) Dibuje el histograma de frecuencias (relativas) de ambos Fondos y comente sus principales similitudes y diferencias.

(b) Realice un contraste de homogeneidad de Kolmogorov-Smirnov. ¿Podemos concluir que ambos Fondos tienen idéntica distribución a lo largo del periodo muestral considerado?

(c) Realice un contraste Chi-cuadrado para discutir si los rendimientos del Fondo A se distribuyen de manera normal con media cero y varianza 1.

(d) A luz de los resultados de los apartados anteriores, sin realizar ningún contraste adicional ¿qué podría afirmar acerca de la distribución de los rendimientos del Fondo B?

A resolver en clase

Ejercicio 15. Sea X los tipos de interés a un año. Tenemos una muestra aleatoria simple con 40 datos. El coeficiente de curtosis es 3,852 y el de asimetría de -0,305. Queremos contrastar si se distribuye normal usando el estadístico de Jarque-Bera. Halle el p-valor de este contraste y comente el valor obtenido en relación a los distintos posible niveles de significación para realizar el contraste.

A resolver en clase

Test. Conteste a las siguientes cuestiones.

1. Un canal de televisión realiza una encuesta a 5000 menores de 30 años, y la siguiente tabla de contingencia resume la información acerca del gusto por las películas de Van Damme:

	gusta	no gusta
chicos	348	3152
chicas	82	1418

el valor del estadístico chi-cuadrado para contrastar la independencia entre el gusto por las películas y el sexo es igual a:

(a) 0.276 (b) 21.26
(c) 26.76 (d) ninguno de los restantes

2. Se pretende contrastar si las subidas del rendimiento a lo largo de un día es la misma para dos activos financieros, A y B. Para esto, se divide el día entre 'por la mañana', 'al mediodía' y 'por la tarde'. La información muestral es la siguiente:

	mañana	mediodía	tarde
A	2400	1900	700
B	3100	2700	1200

(a) Rechazamos la hipótesis nula de homogeneidad al 5 % pero no al 10 %
(b) Aceptamos la hipótesis nula de homogeneidad al 5 %
(c) Rechazamos la hipótesis nula de homogeneidad al 5 % y 10 %
(d) Aceptamos la hipótesis nula de homogeneidad a niveles de significación inferiores al 10 %

3. Queremos contrastar si un dado está trucado o no. Para esto lo lanzamos 300 y apuntamos el número de veces que sale cada valor.

	1	2	3	4	5	6
	44	62	52	45	50	47

El valor del estadístico de Kolmogorov de una muestra y el chi-cuadrado son, respectivamente:

(a) 0,020 y 4,36 (b) 0,020 y 0,436
(c) 0,027 y 4,36 (d) 0,027 y 0,436

4. Algunos profesores opinan que las notas obtenidas no dependen de que el alumno estudie o no el día antes del examen. Con los datos de la siguiente tabla:

	Suspenso	Aprobado	Notable	Sobresaliente
Estudian	16	15	20	40
No estudian	15	40	20	15

¿Qué afirmación sería la correcta? (use un contraste chi-cuadrado):

(a) Las calificaciones son independientes para un nivel de significación del 5 %.
(b) Las calificaciones no son independientes para un nivel de significación del 10 %
(c) No tenemos suficiente información para llevar a cabo el contraste
(d) Este contraste sería equivalente a uno de igualdad de proporciones

5. Se quiere contrastar que un dado no está trucado. Para esto se tira 120 veces y se apunta el número de veces que sale cada número. Se lleva a cabo un contraste de Kolmogorov-Smirnov. La información muestral queda resumida en la siguiente tabla:

valor del dado	1	2	3	4	5	6
numero de veces	21	16	25	18	18	22

El valor del estadístico de contraste es:

(a) 5 (b) 0.025
(c) Ninguna de las anteriores (d) 0.017

6. El p-valor de un determinado contraste es 0.12. ¿Qué nos dice esta información?

(a) Para niveles de significación superiores a 0.12, no aceptamos la hipótesis nula.
(b) Para niveles de significación inferiores a 0.12, no aceptamos la hipótesis nula.
(c) La probabilidad de rechazar la hipótesis nula siendo ésta verdadera es 0.12.
(d) La probabilidad de rechazar la hipótesis nula siendo ésta falsa es 0.12.

Lista de Trasparencias

6. Bibliografía

Novales, A. (1997). *Estadística y Econometría*. McGraw-Hill, Madrid, primera ed. ISBN 84-481-0798-5. 1, 4, 5, 6, 7, 10, 17

Peña, D. (2001). *Fundamentos de Estadística*. Alianza Editorial, Madrid. ISBN 84-206-8696-4. 1, 17

Peña, D. (2002). *Regresión y diseño de experimentos*. Alianza Editorial, Madrid. ISBN 84-206-8695-6. 1

Peña, D. y Romo, J. (1997). *Introducción a la Estadística para la Ciencias Sociales*. McGraw-Hill, Madrid. ISBN 84-481-1617-8. 1

A. Chuletario y Tablas

Contraste de Jarque-Bera $JB = n\left[\frac{AS^2}{6} + \frac{(K-3)^2}{24}\right] \sim \chi^2_{(2)}$

Contraste Chi cuadrado $\sum_{i=1}^{k} \frac{(T_i - O_i)^2}{T_i} \sim \chi^2_{k-1}$, donde T_i y O_i son, respectivamente las i-ésimas frecuencias absolutas esperadas y observadas.

Kolmogorov-Smirnov para una sola muestra: $D_n = \sup |F_n(x) - F(x)|$, donde $F_n(x)$ es la función de distribución empírica (o muestral), y $F(x)$ es la función de distribución de H_0.

	Nivel significación				
Tamaño muestral	0.20	0.15	0.10	0.05	0.01
mas de 35	$\frac{1.07}{\sqrt{n}}$	$\frac{1.14}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

Kolmogorov-Smirnov para dos muestras: $D_n = \sup |F_1(x) - F_2(x)|$. Cuando n_1 y n_2 son grandes; y donde $F_1(x)$ y $F_2(x)$ son funciones de distribución empíricas (o muestrales)

$$\text{Nivel crítico: } D_{\alpha,n_1,n_2} \approx k\sqrt{\frac{n_1+n_2}{n_1 n_2}},$$

donde $k = 1.07;\ 1.22;\ 1.52$; para α igual a 10 %, 5 % y 1 % respectivamente.

Contraste de Wilcoxon. Bajo H_0

$$\mathrm{E}\,(T) = 0, \quad \mathrm{Var}\,(T) = \frac{n(n+1)(2n+1)}{6}; \qquad \mathrm{E}\,(T^+) = \frac{n(n+1)}{4}, \quad \mathrm{Var}\,(T^+) = \frac{n(n+1)(2n+1)}{24}$$

Probabilidad acumulada desde 0 hasta x para $X \sim \chi^2_\nu$

ν	0.1 %	0.5 %	1.0 %	2.5 %	5.0 %	10.0 %	12.5 %	20.0 %	25.0 %	33.3 %	50.0 %
1	0.000	0.000	0.000	0.001	0.004	0.016	0.025	0.064	0.102	0.186	0.455
2	0.002	0.010	0.020	0.051	0.103	0.211	0.267	0.446	0.575	0.811	1.386
3	0.024	0.072	0.115	0.216	0.352	0.584	0.692	1.005	1.213	1.568	2.366
4	0.091	0.207	0.297	0.484	0.711	1.064	1.219	1.649	1.923	2.378	3.357
5	0.210	0.412	0.554	0.831	1.145	1.610	1.808	2.343	2.675	3.216	4.351
6	0.381	0.676	0.872	1.237	1.635	2.204	2.441	3.070	3.455	4.074	5.348
7	0.598	0.989	1.239	1.690	2.167	2.833	3.106	3.822	4.255	4.945	6.346
8	0.857	1.344	1.646	2.180	2.733	3.490	3.797	4.594	5.071	5.826	7.344
9	1.152	1.735	2.088	2.700	3.325	4.168	4.507	5.380	5.899	6.716	8.343
10	1.479	2.156	2.558	3.247	3.940	4.865	5.234	6.179	6.737	7.612	9.342
11	1.834	2.603	3.053	3.816	4.575	5.578	5.975	6.989	7.584	8.514	10.341
12	2.214	3.074	3.571	4.404	5.226	6.304	6.729	7.807	8.438	9.420	11.340
13	2.617	3.565	4.107	5.009	5.892	7.042	7.493	8.634	9.299	10.331	12.340
14	3.041	4.075	4.660	5.629	6.571	7.790	8.266	9.467	10.165	11.245	13.339
15	3.483	4.601	5.229	6.262	7.261	8.547	9.048	10.307	11.037	12.163	14.339
16	3.942	5.142	5.812	6.908	7.962	9.312	9.837	11.152	11.912	13.083	15.338
17	4.416	5.697	6.408	7.564	8.672	10.085	10.633	12.002	12.792	14.006	16.338
18	4.905	6.265	7.015	8.231	9.390	10.865	11.435	12.857	13.675	14.931	17.338
19	5.407	6.844	7.633	8.907	10.117	11.651	12.242	13.716	14.562	15.859	18.338
20	5.921	7.434	8.260	9.591	10.851	12.443	13.055	14.578	15.452	16.788	19.337
21	6.447	8.034	8.897	10.283	11.591	13.240	13.873	15.445	16.344	17.720	20.337
22	6.983	8.643	9.542	10.982	12.338	14.041	14.695	16.314	17.240	18.653	21.337
23	7.529	9.260	10.196	11.689	13.091	14.848	15.521	17.187	18.137	19.587	22.337
24	8.085	9.886	10.856	12.401	13.848	15.659	16.351	18.062	19.037	20.523	23.337
25	8.649	10.520	11.524	13.120	14.611	16.473	17.184	18.940	19.939	21.461	24.337
26	9.222	11.160	12.198	13.844	15.379	17.292	18.021	19.820	20.843	22.399	25.336
27	9.803	11.808	12.879	14.573	16.151	18.114	18.861	20.703	21.749	23.339	26.336
28	10.391	12.461	13.565	15.308	16.928	18.939	19.704	21.588	22.657	24.280	27.336
29	10.986	13.121	14.256	16.047	17.708	19.768	20.550	22.475	23.567	25.222	28.336
30	11.588	13.787	14.953	16.791	18.493	20.599	21.399	23.364	24.478	26.165	29.336
35	14.688	17.192	18.509	20.569	22.465	24.797	25.678	27.836	29.054	30.894	34.336
40	17.916	20.707	22.164	24.433	26.509	29.051	30.008	32.345	33.660	35.643	39.335
45	21.251	24.311	25.901	28.366	30.612	33.350	34.379	36.884	38.291	40.407	44.335
50	24.674	27.991	29.707	32.357	34.764	37.689	38.785	41.449	42.942	45.184	49.335
55	28.173	31.735	33.570	36.398	38.958	42.060	43.220	46.036	47.610	49.972	54.335
60	31.738	35.534	37.485	40.482	43.188	46.459	47.680	50.641	52.294	54.770	59.335

Probabilidad acumulada desde 0 hasta x para $X \sim \chi^2_\nu$

ν	60.0 %	66.7 %	75.0 %	80.0 %	87.5 %	90.0 %	95.0 %	97.5 %	99.0 %	99.5 %	99.9 %
1	0.708	0.936	1.323	1.642	2.354	2.706	3.841	5.024	6.635	7.879	10.828
2	1.833	2.197	2.773	3.219	4.159	4.605	5.991	7.378	9.210	10.597	13.816
3	2.946	3.405	4.108	4.642	5.739	6.251	7.815	9.348	11.345	12.838	16.266
4	4.045	4.579	5.385	5.989	7.214	7.779	9.488	11.143	13.277	14.860	18.467
5	5.132	5.730	6.626	7.289	8.625	9.236	11.070	12.833	15.086	16.750	20.515
6	6.211	6.867	7.841	8.558	9.992	10.645	12.592	14.449	16.812	18.548	22.458
7	7.283	7.992	9.037	9.803	11.326	12.017	14.067	16.013	18.475	20.278	24.322
8	8.351	9.107	10.219	11.030	12.636	13.362	15.507	17.535	20.090	21.955	26.125
9	9.414	10.215	11.389	12.242	13.926	14.684	16.919	19.023	21.666	23.589	27.877
10	10.473	11.317	12.549	13.442	15.198	15.987	18.307	20.483	23.209	25.188	29.588
11	11.530	12.414	13.701	14.631	16.457	17.275	19.675	21.920	24.725	26.757	31.264
12	12.584	13.506	14.845	15.812	17.703	18.549	21.026	23.337	26.217	28.300	32.910
13	13.636	14.595	15.984	16.985	18.939	19.812	22.362	24.736	27.688	29.819	34.528
14	14.685	15.680	17.117	18.151	20.166	21.064	23.685	26.119	29.141	31.319	36.123
15	15.733	16.761	18.245	19.311	21.384	22.307	24.996	27.488	30.578	32.801	37.697
16	16.780	17.840	19.369	20.465	22.595	23.542	26.296	28.845	32.000	34.267	39.252
17	17.824	18.917	20.489	21.615	23.799	24.769	27.587	30.191	33.409	35.718	40.790
18	18.868	19.991	21.605	22.760	24.997	25.989	28.869	31.526	34.805	37.156	42.312
19	19.910	21.063	22.718	23.900	26.189	27.204	30.144	32.852	36.191	38.582	43.820
20	20.951	22.133	23.828	25.038	27.376	28.412	31.410	34.170	37.566	39.997	45.315
21	21.991	23.201	24.935	26.171	28.559	29.615	32.671	35.479	38.932	41.401	46.797
22	23.031	24.268	26.039	27.301	29.737	30.813	33.924	36.781	40.289	42.796	48.268
23	24.069	25.333	27.141	28.429	30.911	32.007	35.172	38.076	41.638	44.181	49.728
24	25.106	26.397	28.241	29.553	32.081	33.196	36.415	39.364	42.980	45.559	51.179
25	26.143	27.459	29.339	30.675	33.247	34.382	37.652	40.646	44.314	46.928	52.620
26	27.179	28.520	30.435	31.795	34.410	35.563	38.885	41.923	45.642	48.290	54.052
27	28.214	29.580	31.528	32.912	35.570	36.741	40.113	43.195	46.963	49.645	55.476
28	29.249	30.639	32.620	34.027	36.727	37.916	41.337	44.461	48.278	50.993	56.892
29	30.283	31.697	33.711	35.139	37.881	39.087	42.557	45.722	49.588	52.336	58.301
30	31.316	32.754	34.800	36.250	39.033	40.256	43.773	46.979	50.892	53.672	59.703
35	36.475	38.024	40.223	41.778	44.753	46.059	49.802	53.203	57.342	60.275	66.619
40	41.622	43.275	45.616	47.269	50.424	51.805	55.758	59.342	63.691	66.766	73.402
45	46.761	48.510	50.985	52.729	56.052	57.505	61.656	65.410	69.957	73.166	80.077
50	51.892	53.733	56.334	58.164	61.647	63.167	67.505	71.420	76.154	79.490	86.661
55	57.016	58.945	61.665	63.577	67.211	68.796	73.311	77.380	82.292	85.749	93.168
60	62.135	64.147	66.981	68.972	72.751	74.397	79.082	83.298	88.379	91.952	99.607

	α				
n	0.2	0.1	0.05	0.02	0.01
1	0.9000	0.9500	0.9750	0.9900	0.9950
2	0.6838	0.7764	0.8419	0.9000	0.9293
3	0.5648	0.6360	0.7076	0.7846	0.8290
4	0.4927	0.5652	0.6239	0.6889	0.7342
5	0.4470	0.5094	0.5633	0.6272	0.6685
6	0.4104	0.4680	0.5193	0.5774	0.6166
7	0.3815	0.4361	0.4834	0.5384	0.5758
8	0.3583	0.4096	0.4543	0.5065	0.5418
9	0.3391	0.3875	0.4300	0.4796	0.5133
10	0.3226	0.3687	0.4092	0.4566	0.4889
11	0.3083	0.3524	0.3912	0.4367	0.4677
12	0.2958	0.3382	0.3754	0.4192	0.4490
13	0.2847	0.3255	0.3614	0.4036	0.4325
14	0.2748	0.3142	0.3489	0.3897	0.4176
15	0.2659	0.3040	0.3376	0.3771	0.4042
16	0.2578	0.2947	0.3273	0.3657	0.3920
17	0.2504	0.2863	0.3180	0.3553	0.3809
18	0.2436	0.2785	0.3094	0.3457	0.3706
19	0.2373	0.2714	0.3014	0.3369	0.3612
20	0.2316	0.2647	0.2941	0.3287	0.3524
21	0.2262	0.2586	0.2872	0.3210	0.3443
22	0.2212	0.2528	0.2809	0.3139	0.3367
23	0.2165	0.2475	0.2749	0.3073	0.3295
24	0.2120	0.2424	0.2693	0.3010	0.3229
25	0.2079	0.2377	0.2640	0.2952	0.3166
26	0.2040	0.2332	0.2591	0.2896	0.3106
27	0.2003	0.2290	0.2544	0.2844	0.3050
28	0.1968	0.2250	0.2499	0.2794	0.2997
29	0.1935	0.2212	0.2457	0.2747	0.2947
30	0.1903	0.2176	0.2417	0.2702	0.2899
31	0.1873	0.2141	0.2379	0.2660	0.2853
32	0.1844	0.2108	0.2342	0.2619	0.2809
33	0.1817	0.2077	0.2308	0.2580	0.2768
34	0.1791	0.2047	0.2274	0.2543	0.2728
35	0.1766	0.2018	0.2242	0.2507	0.2690
36	0.1742	0.1991	0.2212	0.2473	0.2653
37	0.1719	0.1965	0.2183	0.2440	0.2618
38	0.1697	0.1939	0.2154	0.2409	0.2584
39	0.1675	0.1915	0.2127	0.2379	0.2552
40	0.1655	0.1891	0.2101	0.2349	0.2521
> 40	$1.07/\sqrt{n}$	$1.22/\sqrt{n}$	$1.36/\sqrt{n}$	$1.52/\sqrt{n}$	$1.63/\sqrt{n}$

Cuadro 1: Kolmogorov-Smirnov de una muestra

	Nominal α (valores para una cola entre paréntesis)					
	0.20	0.10	0.05	0.02	0.02	0.002
n	(0.10)	(0.05)	(0.025)	(0.01)	(0.005)	(0.001)
4	1.000	1.000	-	-	-	-
5	0.800	0.900	1.000	1.000	-	-
6	0.657	0.829	0.886	0.943	1.000	-
7	0.571	0.714	0.786	0.893	0.929	1.000
8	0.524	0.643	0.738	0.833	0.881	0.952
9	0.483	0.600	0.700	0.783	0.833	0.917
10	0.455	0.564	0.648	0.745	0.794	0.879
11	0.427	0.536	0.618	0.709	0.755	0.845
12	0.406	0.503	0.587	0.678	0.727	0.818
13	0.385	0.484	0.560	0.648	0.703	0.791
14	0.367	0.464	0.538	0.626	0.679	0.771
15	0.354	0.446	0.521	0.604	0.654	0.750
16	0.341	0.429	0.503	0.582	0.635	0.729
17	0.328	0.414	0.488	0.566	0.618	0.711
18	0.317	0.401	0.472	0.550	0.600	0.692
19	0.309	0.391	0.460	0.535	0.584	0.675
20	0.299	0.380	0.447	0.522	0.570	0.662
21	0.292	0.370	0.436	0.509	0.556	0.647
22	0.284	0.361	0.425	0.497	0.544	0.633
23	0.278	0.353	0.416	0.486	0.532	0.621
24	0.271	0.344	0.407	0.476	0.521	0.609
25	0.265	0.337	0.398	0.466	0.511	0.597
26	0.259	0.331	0.390	0.457	0.501	0.586
27	0.255	0.324	0.383	0.449	0.492	0.576
28	0.250	0.318	0.375	0.441	0.483	0.567
29	0.245	0.312	0.368	0.433	0.475	0.558
30	0.240	0.306	0.362	0.425	0.467	0.549
31	0.236	0.301	0.356	0.419	0.459	0.540
32	0.232	0.296	0.350	0.412	0.452	0.532
33	0.229	0.291	0.345	0.405	0.446	0.525
34	0.225	0.287	0.340	0.400	0.439	0.517
35	0.222	0.283	0.335	0.394	0.433	0.510
36	0.219	0.279	0.330	0.388	0.427	0.503
37	0.215	0.275	0.325	0.383	0.421	0.497
38	0.212	0.271	0.321	0.378	0.415	0.491
39	0.210	0.267	0.317	0.373	0.410	0.485
40	0.207	0.264	0.313	0.368	0.405	0.479
41	0.204	0.261	0.309	0.364	0.400	0.473
42	0.202	0.257	0.305	0.359	0.396	0.468
43	0.199	0.254	0.301	0.355	0.391	0.462
44	0.197	0.251	0.298	0.351	0.386	0.457
45	0.194	0.248	0.294	0.347	0.382	0.452
46	0.192	0.246	0.291	0.343	0.378	0.448
47	0.190	0.243	0.288	0.340	0.374	0.443
48	0.188	0.240	0.285	0.336	0.370	0.439
49	0.186	0.238	0.282	0.333	0.366	0.434
50	0.184	0.235	0.279	0.329	0.363	0.430
51	0.182	0.233	0.276	0.326	0.359	0.426
52	0.180	0.231	0.274	0.323	0.356	0.422
53	0.179	0.228	0.271	0.320	0.352	0.418
54	0.177	0.226	0.268	0.317	0.349	0.414
55	0.175	0.224	0.266	0.314	0.346	0.411
56	0.174	0.222	0.264	0.311	0.343	0.407
57	0.172	0.220	0.261	0.308	0.340	0.404
58	0.171	0.218	0.259	0.306	0.337	0.400
59	0.169	0.216	0.257	0.303	0.334	0.397

Cuadro 2: Valores críticos para el coeficiente de **correlación de rangos de Spearman**.

B. Relación entre la distribución de Poisson y la Binomial.

Sea $X \sim \text{Poisson}\,(\lambda)$ una v.a. que mide el número de sucesos en un periodo fijo T (por ejemplo Nº de llamadas recibidas en 10 horas). Podemos dividir el tiempo total T en intervalos cada vez más pequeños (10 periodos de una hora, 600 periodos de un minuto, 36000 periodos de un segundo, …). Entonces podemos considerar X como una binomial donde p es la probabilidad del suceso en cada uno de los intervalos, y el número de repeticiones $n = 10$ en el primer caso, $n = 600$ en el segundo caso, …; donde

$$\mathrm{E}\,(X) = np$$

si llamo $\lambda = np$, entonces $p = \lambda/n$.

Por tanto, podemos aproximar una binomial a una poisson del siguiente modo:

$$P_X\,(x) = \binom{n}{x}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}$$

y tomando límites

$$\lim_{n\to\infty} P_X\,(x) = \frac{\lambda^x}{x!}\lim_{n\to\infty}\frac{n(n-1)\cdots(n+x+1)}{\left(1-\frac{\lambda}{n}\right)^x n^x}\left(1-\frac{\lambda}{n}\right)^n$$

donde

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$$

y

$$\lim_{n\to\infty} = \frac{n}{(n-x)}\cdot\frac{(n-1)}{(n-\lambda)}\cdots\frac{(n-x+1)}{(n-\lambda)} = 1;$$

Así pues, en el límite

$$P_X\,(x) = \frac{\lambda^x}{x!}e^{-\lambda} \qquad x = 0, 1, 2, \ldots$$

que es la función de cuantía de una Poisson (Demostración de Peña, 2001, pp. 172) (Véase también Novales, 1997, pp.210).

Partes del temario

- Tema 1 IntEctr-T01
- Tema 2 IntEctr-T02
- Tema 3 IntEctr-T03
- Tema 4 IntEctr-T04
- Tema 5 IntEctr-T05
- Tema 6 IntEctr-T06
- Tema 7 IntEctr-T07

Soluciones a los Ejercicios

Ejercicio 1. El valor del estadístico del contraste es: $\frac{(750-680)^2}{750} + \frac{(250-320)^2}{250} = 26.13$, que excede con mucho el valor crítico de las tablas correspondientes a una chi-cuadrado de un grado de libertad, por lo que rechazamos la hipótesis nula.

Ejercicio 1

Ejercicio 2(a)

1. **Jarque-Bera:** $RC = \{\boldsymbol{x}$ tales que: $JB > 4.605\}$
$$100 \cdot \left(\frac{-0.24^2}{6} + \frac{0.23^2}{24} \right) = 1.18 < 4.605 = \chi^2_{2,\ 90\,\%}$$
Por lo que no rechazamos H_0 de normalidad.

2. **Kolmogorov-Smirnov:** Primero calculo la probabilidad de cada intervalo bajo H_0 empleando las tablas de la $\mathrm{N}\,(0\,,\,1)$:
   - $a = P\,(x > 1.5) = 1 - .9332 = 0.0668$
   - $b = P\,(1 < x \le 1.5) = (1 - .8413) - a = 0.0919$
   - $c = P\,(0.5 < x \le 1) = (1 - .6915) - a - b = 0.1498$
   - $d = P\,(-0.5 < x \le 0.5) = 2 \cdot P\,(0 > x \le 0.5) = 2 * (.6915 - .5) = 0.383$

   Y puesto que la función de densidad de una Normal es simétrica, no es necesario calcular más probabilidades:
   Tabla probabilidades teóricas bajo H_0:

Intervalos	< -1.5	$[-1.5; -1)$	$[-1; -0,5)$	$[-0,5; 0,5]$	$(0,5; 1]$	$(1; 1.5]$	> 1.5
prob.$_{H_0}$	0.0668	0.0919	0.1498	0.383	0.1498	0.0919	0.0668

   Acumulando las probabilidades teóricas obtenemos la función de distribución teórica bajo H_0, $F_z(z)$; y para calcular la función de distribución empírica, $F_n(x)$, debemos dividir el número de días de la tabla del enunciado por 100, y acumular

z	-1.5	-1	$-0,5$	$0,5$	1	1.5	∞
$F_z(z)$	0.0668	0.1587	0.3085	0.6915	0.8413	0.9332	1
$F_n(x)$	0.0900	0.1500	0.2700	0.6800	0.8700	0.9500	1
D_n = Max diff			0.0385				

   Y puesto que $0.0385 < \frac{1.22}{\sqrt{100}} = 0.122$ no podemos rechazar H_0 de distribución $\mathrm{N}\,(0\,,\,1)$ al 10 % de significación.

3. **Chi cuadrado:** Sólo queda calcular las frecuencias absolutas para poder calcular este contraste; que da:
$$\frac{(9\ -\ 6.68)^2}{6.68} + \frac{(6\ -\ 9.19)^2}{9.19} + \frac{(12\ -\ 14.98)^2}{14.98} + \cdots + \frac{(5\ -\ 6.68)^2}{6.68} = 4.3516$$
Puesto que este valor es menor que lo que aparece en las tablas de la Chi cuadrado con *seis grados de libertad* (10.64) no podemos rechazar H_0 con un nivel de significación del 10 %.

$\square$

Ejercicio 2(b) Puesto que en el apartado anterior no hemos rechazado la hipótesis de distribución normal, podemos emplear los estadísticos basados en distribución normal.

En este caso podemos emplear el estadístico: $\frac{\bar{x}-\mu}{\sqrt{s^2/n}} \sim t_{n-1}$

1. $H_0 : \mu = 1; \qquad H_1 : \mu > 1$
2. $RC = \left\{ \boldsymbol{x} \text{ tales que: } \frac{\bar{x}-1}{\sqrt{s^2/100}} > t_{99,\ \alpha} \right\}$
3. Por una parte, $t_{99,\ 5\,\%} = 1.645$, y $t_{99,\ 10\,\%} = 1.282$; por otra $\frac{1.15-1}{\sqrt{0.04/100}} = 7.5$ que supera ambos niveles críticos. Por tanto se rechaza H_0 tanto al 10 % como al 5 % de nivel de significción.

$\square$

Ejercicio 2(c)

- El valor 7.5 es tan elevado, que no aparece en las tablas por lo que podemos afirmar que el p-valor es aproximadamente cero.
- Es la probabilidad, calculada bajo H_0, de obtener un valor numérico que aporte más evidencia en contra de H_0, que la aportada por el estadístico que hemos obtenido (en nuestro caso que 7.5).

- Cuanto menor es el p-valor, menor ha de ser el nivel de significación (menor que el p-valor) para poder aceptar la hipótesis nula.

□

Ejercicio 2(d) Si hubiésemos rechazado la hipótesis de normalidad, no habríamos podido emplear el estadístico $\frac{\bar{x}-\mu}{\sqrt{s^2/n}}$ que se basa en el supuesto de normalidad en la distribución.

□

Ejercicio 3.

z	1/3	2/3	1
$F_z(z)$	1/3	2/3	3/3
$F_n(x)$	4/10	5/10	1
D_n = Max diff	1/15	**1/6**	0

Puesto que $1/6 = 0.16$ es un nº menor que cualquiera de los que vienen en la tabla para un tamaño muestral de 10 datos, no es posible rechazar la hipótesis de distribución Uniforme $(0,\ 1)$ con una significación del 20 % o menos.

Ejercicio 3

Ejercicio 4.

$$n\left[\frac{\widehat{AS}^2}{6}+\frac{(\widehat{K}-3)^2}{24}\right]=40\cdot\left[\frac{0.5^2}{6}+\frac{(2.5-3)^2}{24}\right]=2.08$$

Por tanto, el p-valor es aproximadamente un 0.35.

La conclusión es que se rechazaría la hipótesis de distribución normal de la muestra con niveles de significación superiores al 35 %.

Ejercicio 4

Ejercicio 5. H_0: idéntica distribución; Nº de alumnos en 2005: 500; Nº de alumnos en 2006: 700;

Probabilidades estimadas (bajo H_0): $P\,(\text{suspenso}) = \frac{240+310}{1200} = 0.458$; $P\,(\text{aprobado}) = 0.384$; $P\,(\text{nota}) = 0.158$

Tabla de frecuencias absolutas esperadas bajo H_0:

	suspensos	aprobados	con nota
2005	0.458×500 =229.1	191.7	79.2
2006	320.8	268.4	110.8

El estadístico se distribuye como una Chi cuadrado con dos grados de libertad y

$$\sum_{i=1}^{4}\frac{(T_i-O_i)^2}{T_i}=2.74;$$

por tanto el p-valor es aproximadamente 0.25; esto quiere que se rechaza H_0 para niveles de significación superiores al 25 %. Así que al 5 % y al 10 % no se rechaza H_0.

Ejercicio 5

Ejercicio 6. Las funciones de distribución empíricas son:

	1	2	3
2005	0.48	0.86	1
2006	0.44	0.82	1
diferencia	0.04	0.04	0

Por lo tanto 0.04

Ejercicio 6

Ejercicio 8.

Kolmogorov-Smirnov Calculando la suma acumulada para el caso de Madrid, es decir:

Madrid	-2	-1	0	1	2
suma acumulada	100	250	375	475	500

y dividiendo por el numero de observaciones (500 para Madrid), obtenemos la función de distribución empírica para Madrid.

Procediendo de manera similar en el caso de Barcelona, obtenemos las siguientes func. de distribución empíricas y las diferencias entre ellas:

Func. dist	-2	-1	0	1	2
$F_{Madr}(x)$	0.20	0.50	0.75	0.95	1
$F_{Barc}(x)$	0.15	0.50	0.75	0.90	1
diferencias	0.05	0.00	0.00	0.05	0

Así pues el estadístico Kolmogorov-Smirnov es la máxima diferencia entre ambas funciones de distribución $D_{n,m} = \sup |F_{Madr}(x) - F_{Barc}(x)| = 0.05$.

Los valores críticos para niveles de significación del 1 % y 10 % son respectivamente

- $\alpha_{0.01,\ 500,\ 450} = 1.52 \cdot \sqrt{\frac{500+450}{500 \cdot 450}} = 0.098$
- $\alpha_{0.1,\ 500,\ 450} = 1.07 \cdot \sqrt{\frac{500+450}{500 \cdot 450}} = 0.069$

Así pues no rechazamos H_0 ni al 1 % ni al 10 % de nivel de significación.

Chi cuadrado Las frecuencias teóricas son las siguientes:

	-2	-1	0	1	2
Madrid	88.4	161.6	124.7	88.4	36.8
Barcelona	79.6	145.4	112.3	79.6	33.2

La realización del estadístico Chi-cuadrado para este contraste es 16.163, que tenemos que comparar con los valores tabulados de una chi-cuadrado de 4 grados de libertad al 1 % y 10 % de significación. Estos valores son 13,28 y 7,78, respectivamente. En ambos casos, por tanto, rechazamos la hipótesis nula de homogeneidad de muestras. Nótese que la respuesta es la contraria que en el caso del Kolmogorov. En el caso del Chi-cuadrado, si la confianza fuese del 99,9 %, no rechazaríamos la nula de homogeneidad, por ejemplo.

Ejercicio 8

Ejercicio 9.

$$P\,(\text{ser chica}) = \frac{82+1418}{5000} = 0.3 \qquad P\,(\text{ser chico}) = 0.7$$

$$P\,(\text{gusta}) = \frac{348+82}{5000} = \frac{43}{500} \qquad P\,(\text{no gusta}) = \frac{457}{500}$$

Frecuencias teóricas:

	gusta	no gusta
chicos	$0.7 \times \frac{457}{500} \times 5000 = 301$	3199
chicas	129	1371

$\sum_{i=1}^{k} \frac{(T_i - O_i)^2}{T_i} = 26.76$; que es un valor muy grande para una χ^2 con un grado de libertad, por lo que rechazamos el contraste a casi cualquier nivel de significación (p-valor es casi cero) .

Ejercicio 9

Ejercicio 10. Que no se rechaza la hipótesis nula planteada en el contraste con niveles de significación de 12 % o menos.

Ejercicio 10

Ejercicio 11. Debemos estimar las probabilidades teóricas de ser hombre o mujer, y votar centro-derecha o centro-izquierda, sabiendo que la población total son 270 individuos.
$P\,(\text{H}) = (75+80)/270 = 0.574$; por lo que $P\,(\text{M}) = 1 - P\,(\text{H}) = 0.426$
$P\,(\text{C-D}) = (50+75)/270 = 0.463$; por lo que $P\,(\text{C-I}) = 1 - P\,(\text{C-D}) = 0.537$
Si ambas variables son independientes, en teoría la función de cuantía conjunta debería ser igual al producto de las marginales (y la tabla teórica igual a la función de cuantía conjunta multiplicada por 270)

X/Y	mujer	hombre
centro-derechas	$P\,(\text{M}) \cdot P\,(\text{C-D}) \cdot 270 = 53.254$	$P\,(\text{H}) \cdot P\,(\text{C-D}) \cdot 270 = 71.756$
centro-izquierdas	$P\,(\text{M}) \cdot P\,(\text{C-I}) \cdot 270 = 61.766$	$P\,(\text{H}) \cdot P\,(\text{C-I}) \cdot 270 = 83.224$

Ejercicio 11

Ejercicio 12.

	derecha	izquierda
< 18 mil euros	$\frac{190}{400}\frac{205}{400}400 = 97.4$	$\frac{190}{400}\frac{195}{400}400 = 92.6$
≥18 mil euros	$\frac{210}{400}\frac{205}{400}400 = 107.6$	$\frac{210}{400}\frac{195}{400}400 = 102.4$

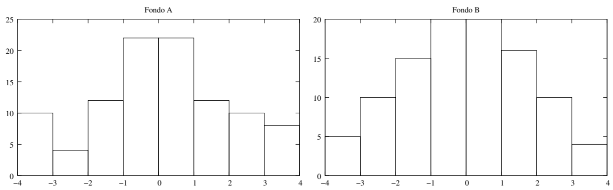
Ejercicio 12

Ejercicio 13.

	Aprobado	suspenso
Mañana	148.58	201.41
Tarde	76.41	103.58

Ejercicio 13

Ejercicio 14(a)



A la vista de los histogramas, se aprecia que el fondo B tiene una distribución simétrica y menos dispersa que la del fondo A (pero esto es una observación sobre la muestra; en realidad desconocemos las características de la distribución teórica de la que provienen los datos).

□

Ejercicio 14(b) Las funciones de distribución empíricas y sus diferencias son:

	< -3	< -2	< -1	< 0	< 1	< 2	< 3	$< \infty$
F_A	0.1	0.14	0.26	0.48	0.7	0.82	0.92	1
F_B	0.05	0.15	0.30	0.50	0.7	0.86	0.96	1
Dif.	0.05	0.01	0.04	0.02	0.0	0.04	0.04	0

Por tanto, el estadístico $D = \text{máx}\,|F_A - F_B|$ toma el valor 0.05

El nivel crítico (para un nivel de significación del 5 %) según las tablas es:

$$1.22 \cdot \sqrt{\frac{200}{10000}} = 0.1725$$

Así pues, con un nivel de significación del 5 % NO podemos rechazar H_0: *idéntica distribución.*

□

Ejercicio 14(c) Necesitamos las frecuencias teóricas u esperadas bajo H_0; que podemos calcular a partir de las tablas de la Normal del siguiente modo

- La primera celda es $\big(P\,(Z < -3)\,\big) \cdot 100 = \big(1 - P\,(Z < 3)\,\big) \cdot 100 = 0.13$
- La segunda es $\Big[\big(P\,(Z < -2)\,\big) - \big(P\,(Z < -3)\,\big)\Big] \cdot 100 = 2.15$
- La tercera es $\Big[\big(P\,(Z < -1)\,\big) - \big(P\,(Z < -2)\,\big)\Big] \cdot 100 = 13.59$
- La cuarta es $\Big[\big(P\,(Z < 0)\,\big) - \big(P\,(Z < -1)\,\big)\Big] \cdot 100 = 34.13$
- ...
- La última es $\big(P\,(Z > 3)\,\big) \cdot 100 = \big(1 - P\,(Z < 3)\,\big) \cdot 100 = 0.13$

donde las frecuencias que faltan son iguales a las ya calculadas por ser la distribución N(0, 1) simétrica respecto al cero.

	< -3	$[-3; -2)$	$[-2; -1)$	$[-1; 0)$	$[0; 1)$	$[1; 2)$	$[2; 3)$	> 3
T_i	0.13	2.15	13.59	34.13	34.13	13.59	2.15	0.13

Por tanto

$$\chi^2 = \sum_{i=1}^{6} \frac{(O_i - T_i)^2}{T_i} = \frac{(10 - 0.13)^2}{0.13} + \frac{(4 - 2.15)^2}{2.15} + \ldots$$

que es claramente superior a 14.1, valor tabulado de una χ^2 con 8-1=7 grados de libertad para un nivel de significación del 5 %; por lo que rechazamos H_0: *distribución normal* con dicho nivel de significación.

□

Ejercicio 14(d) Podemos aventurarnos a suponer que el fondo B es una muestra que tampoco proviene de una distribución normal.

□

Ejercicio 15. Para la muestra disponible $JB = 40\left[\frac{(-.305)^2}{6} + \frac{((3.852)-3)^2}{24}\right] = 1.83$

Así pues, el p-valor de esta muestra es la probabilidad de observar un estadístico JB mayor o igual a 1.83 bajo la hipótesis nula de normalidad, es decir, la probabilidad de que una variable aleatoria con distribución $\chi^2_{(2)}$ sea mayor o igual a 1.83:

$$P\,(JB \geq 1.83) = 0.40$$

Este es un p-valor elevado, esto quiere decir que para niveles de significación inferiores al 40 % (normalmente se emplean niveles del 10 % o menos) no se puede rechazar H_0.

Ejercicio 15

Soluciones a los Tests

Solución al Test: La probabilidad estimada de vivir en el norte es $P_N = 55/100 = 0.55$ Del mismo modo

$$P_S = 45/100 = 0.45;\ P_1 = 30/100 = 0.3;\ P_2 = 0.7;$$

por tanto, bajo H_0 la distribución teórica debería ser

	del equipo 1	del equipo 2	total
zona N	$P_N \cdot P_1 \cdot 100$	$P_N \cdot P_2 \cdot 100$	55
zona S	$P_S \cdot P_1 \cdot 100$	$P_S \cdot P_2 \cdot 100$	45
total	30	70	100

es decir,

	del equipo 1	del equipo 2	total
zona N	16.5	38.5	55
zona S	13.5	31.5	45
total	30	70	100

Fin 1

Solución al Test: El número de grados de libertad del estadístico es $(n-1)(m-1) = 1$

$$\begin{aligned} g(\mathbf{x}) &= \tfrac{(25-16.5)^2}{16.5} + \tfrac{(30-38.5)^2}{38.5} + \tfrac{(5-13.5)^2}{13.5} + \tfrac{(40-31.5)^2}{31.5} \\ &= 13.9 > \chi^2_{1,0.95} = 3.84 \end{aligned}$$

Por tanto, se rechaza con un nivel de significación del 5 %.

Fin 2