

Introducció

de la Teledetecció i els SIG és freqüent realitzar anàlisis estadístics. En aquest context hi ha alguns temes que apareixen de manera recurrent:

anàlisis cal assumir que les variables es distribueixen de forma normal. Així mateix, convé fer algunes reflexions al respecte: **Caracterització de les variables** i **Verificació de normalitat**.

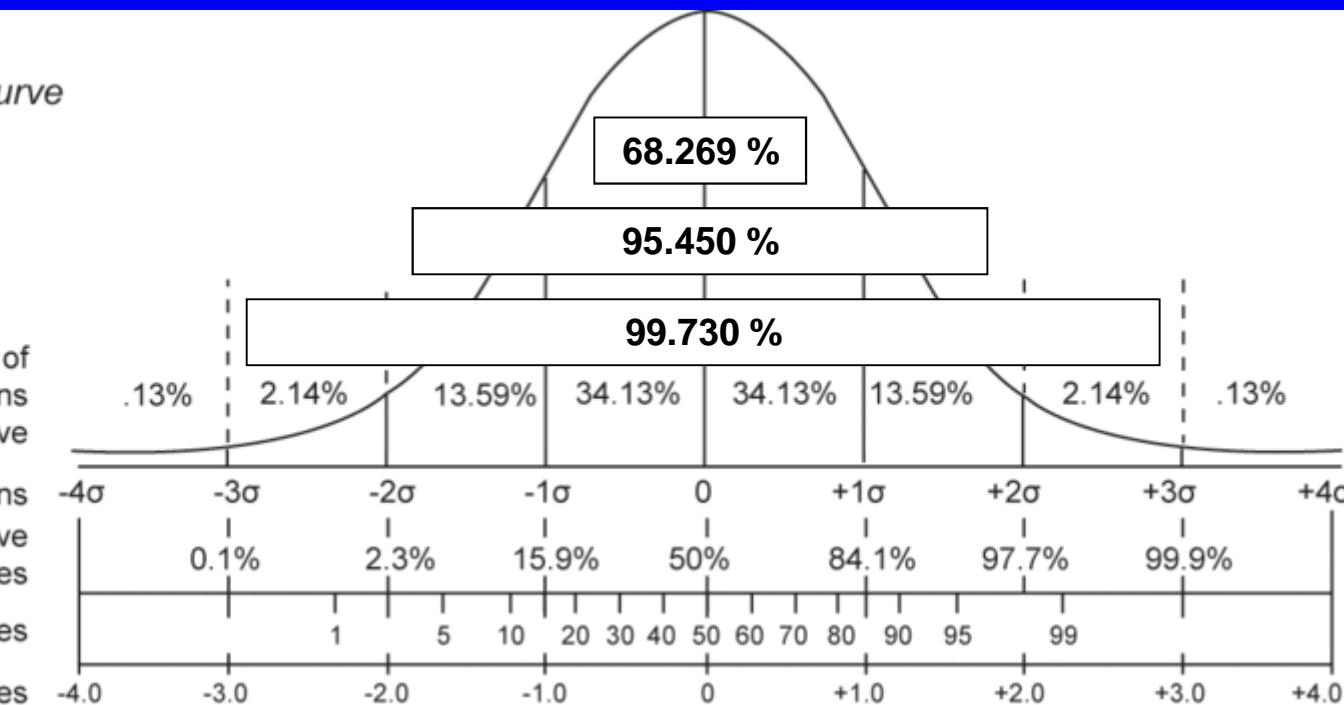
anàlisis cal assumir que no existeix una correlació elevada entre les variables independents: **Correlació en imatges de teledetecció**.

anàlisis pot ser útil reduir el nombre de variables independents i obtenir un subconjunt menor però amb similar poder informatiu: **Anàlisi de components Principals**.

anàlisis cal fer **càlculs de Distàncies estadístiques**.

anàlisis cal tenir present l'**Autocorrelació espacial**.

Caracterizar distribuciones. Recordatorio de la normal univariante



ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

http://en.wikipedia.org/wiki/Standard_score, modificada

1. Generitzar distribucions

En un full MS-Excel en blanc i genereu una columna de valors aleatoris en el rang [-5,5]. Mostreu-los amb 4

• Invoqueu la funció d'Excel que genera un valor aleatori: RANDEGIORIO(). Com que F9 (o reobrir el fitxer, etc) refà la sèrie, convé copiar les fórmules en valors (copiar+enganxat especial).

• Aproveueu visualment, mitjançant un histograma, que la sèrie és realment aleatòria i no segueix, per exemple, una

• Escriviu, en una columna nova, els rangs de dades: des de -5.5 a 5.5, salts de valor 1.0. Després aneu a "Herramientas | Análisis de datos | Histograma" (si no hi fos, afegiu-la amb "Herramientas | Elementos..."). Segons la versió d'Excel, indiqueu "Crear gráfico", "En una hoja nueva"

Peritzar distribucions

Informatiu, calculeu N, mínim, màxim, mitjana i desviació d'aquesta població.

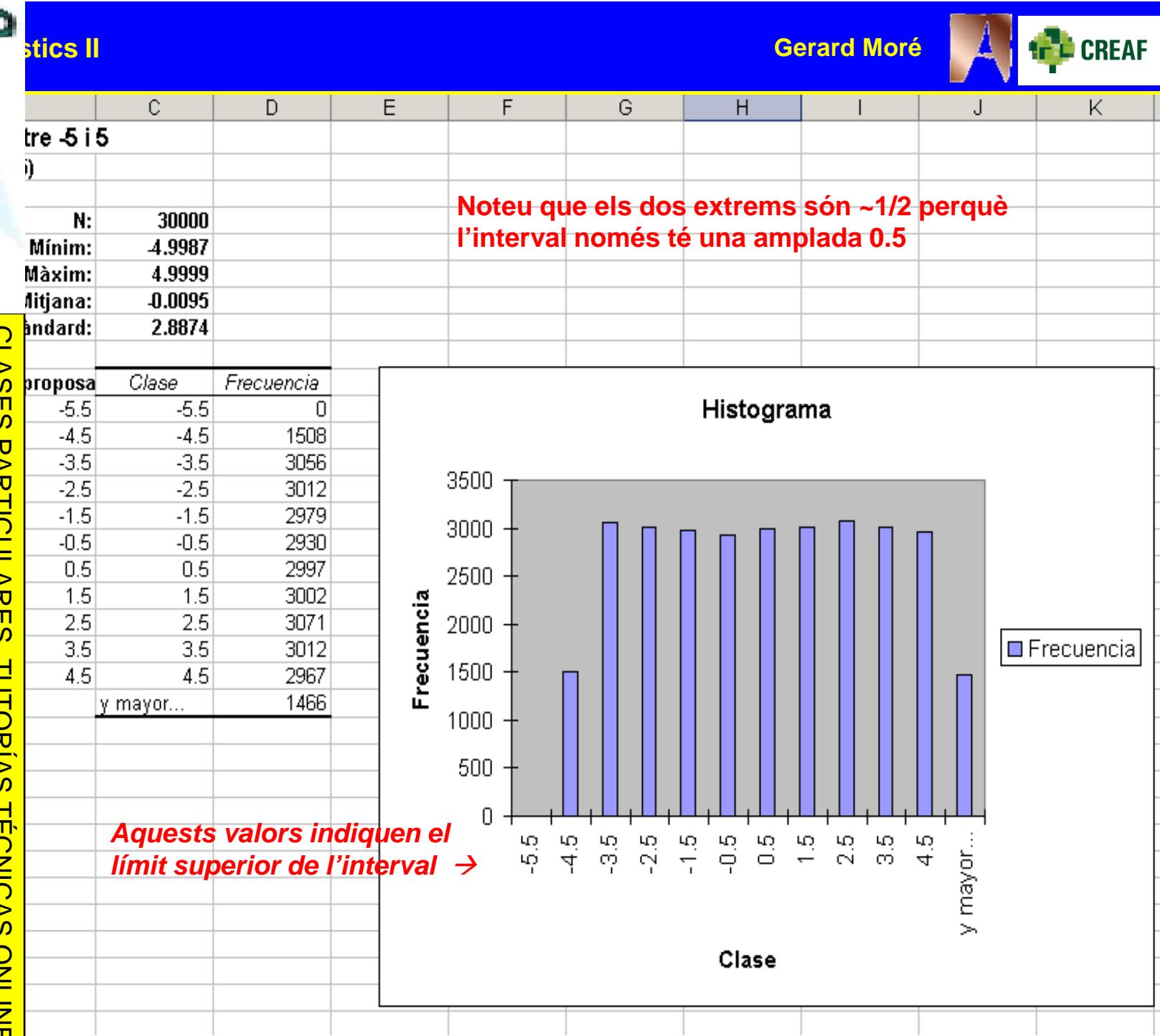
CL 'Contar(), Min(), Max(), Promedio() i DesvEstP() [divideix entre n].

**CLASES PARTICULARES, TUTORIAS TECNICAS ONLINE
LLAMA O ENVIA WHATSAPP: 689 45 44 70**

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70



CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70
...
ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70



s de l'aplicació AleatNormal_AmbParametres.exe, 30000 valors normalment distribuïts al voltant d'una mitjana i una desviació estàndard 1.5. Inclogueu-los en un full de cua en blanc com heu fet en l'exercici anterior.

Redireccioneu la sortida del programa a un fitxer de text, obriu-lo i, després del portapapers, copieu els valors a Excel:

```
AleatNormal_AmbParametres 0 1.5 30000 4 > resu.txt
```

Observeu visualment, mitjançant un histograma, que la mostra segueix realment una normal.

Calculeu N, mínim, màxim, mitjana i desviació estàndard de la mostra imaginària.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70



mitjana 0 i desviació estàndard 1.5

parametres 0 1.5 30000 4 > resu.txt

N: 30000

Mínim: 5.5962

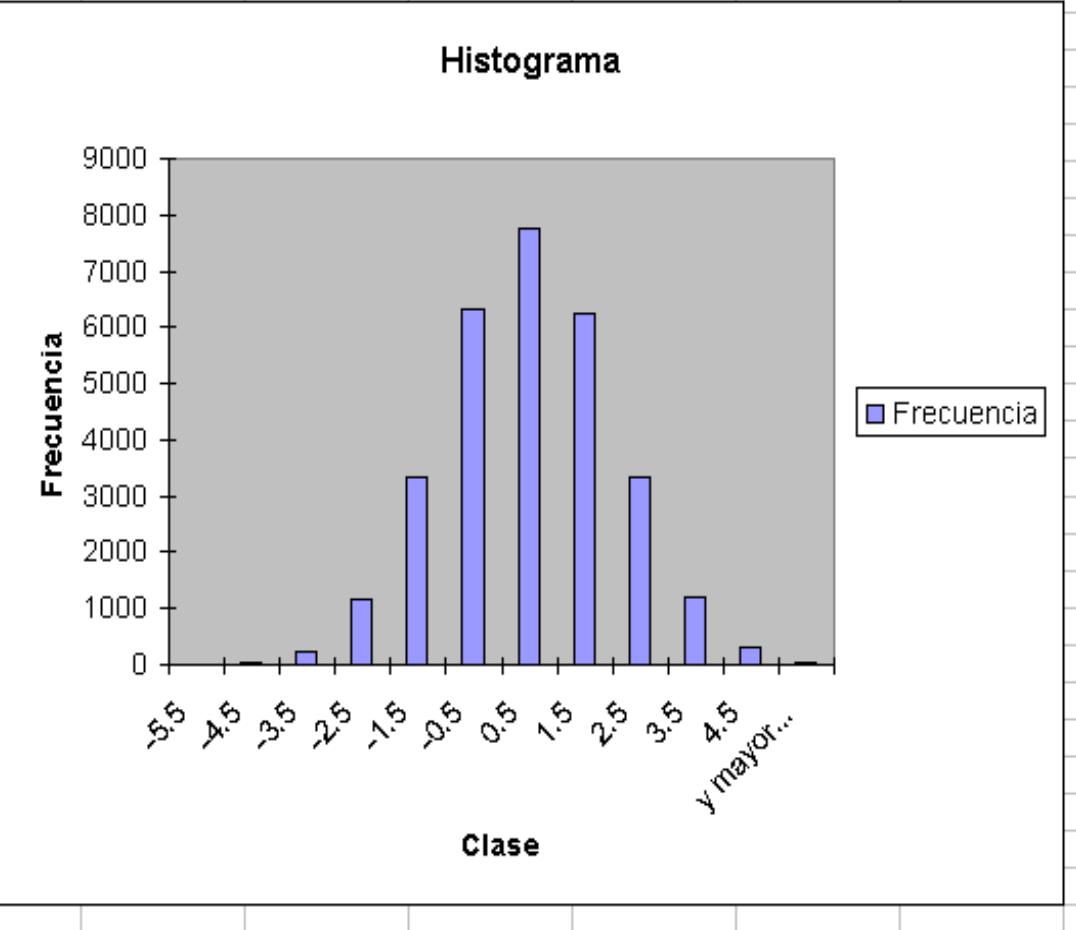
Màxim: 6.9732

mitjana: 0.0079

standard: 1.5107

proposta Clase Frecuencia

-5.5	-5.5	1
-4.5	-4.5	43
-3.5	-3.5	246
-2.5	-2.5	1156
-1.5	-1.5	3325
-0.5	-0.5	6331
0.5	0.5	7765
1.5	1.5	6253
2.5	2.5	3352
3.5	3.5	1198
4.5	4.5	296
y mayor...		34



CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

la normalitat de les dues distribucions anteriors

en les k freqüències dels histogrames calculats (12 en els exemples).

arem en els problemes 7.33 i 12.13 del llibre d'Spiegel "estadística" [McGraw Hill, 2^a Ed.]. La idea és comparar el semblança entre la freqüència ABSOLUTA de cada columna d'una taula i la que tindria una normal teòrica amb la mitjana i desviació estàndard que es dedueix de la mostra que tenim.

voqueu la funció d'Excel que calcula l'àrea acumulada des de $-\infty$ fins al mitjàt: DISTR.NORM(Valor; Mitjana; DesvEst.;1).

final sobre si acceptem o no que s'ajusta a una normal la som fent un test de χ^2 . A tal efecte, recordeu que:

Planteja que la hipòtesi nul·la, H_0 , és que les freqüències observades i esperades no difereixen significativament.

$\chi^2=0$ vol dir que les freqüències observades i les teòriques coincideixen completament

Nombre de graus de llibertat, μ , val $k-1-m$, on m és el nombre de paràmetres que hem estimat de la població (2 en normal). k és el nombre de mostres.

Valors crítics (de tall) depenen dels graus de llibertat i del nivell de significació⁽¹⁾, α , desitjat (compte perquè hi ha taules dels típics nivells de significació 0.05 i 0.01 cal buscar-los a $\chi^2_{.95}$ i $\chi^2_{.99}$).

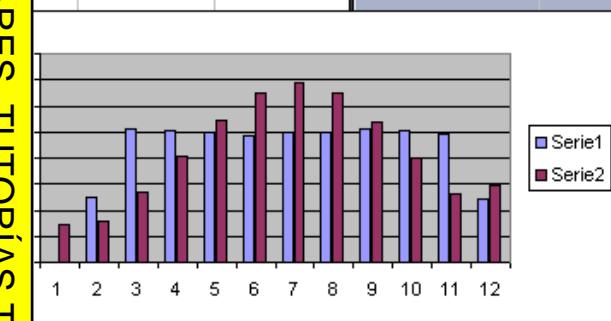
:CONTAR(); $\chi^2_{.95}=\text{PRUEBA.CHI.INV}(0.05; k-1-2)$

Riskat de cometre un error de Tipus I (rebutjar H_0 quan era certa). prob. eT1 però prob. eT2 (acceptar H_0 falsa)

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

	C	D	E	F	G	H
-5 i 5						
N:	30000					
rim:	4.9987					
rim:	4.9999					
ma:	-0.0095					
ard:	2.8874					
Frec. observada Freq. esperada $(\text{Obs-Esp})^2/\text{Esp}$						
Clase (%) (%)						
-5.5 -5.5 0	0.0000	2.8617	2.8617			
-4.5 -4.5 1508	5.0267	3.1334	1.1440			
-3.5 -3.5 3056	10.1867	5.3408	4.3968			
-2.5 -2.5 3012	10.0400	8.0840	0.4733			
-1.5 -1.5 2979	9.9300	10.8659	0.0806			
-0.5 -0.5 2930	9.7667	12.9699	0.7911			
0.5 0.5 2997	9.9900	13.7478	1.0271			
1.5 1.5 3002	10.0067	12.9406	0.6652			
2.5 2.5 3071	10.2367	10.8169	0.0311			
3.5 3.5 3012	10.0400	8.0293	0.5035			
4.5 4.5 2967	9.8900	5.2927	3.9932			
y mayor... 1466	4.8867	5.9170	0.1794			



Exigim un 5% o un 1% de probabilitat de rebutjar H_0 i que fos certa, no ens atrevim a dir. En canvi, podem dir que no és normal acceptant un 10% de probabilitats que ho sigui.

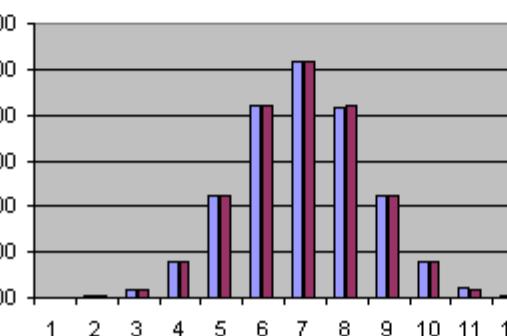
Hi contribueixen les dues freqüències “tallades” a $\frac{1}{2}$ als dos extrems (que simulen les)
i que fins i tot es comparen amb la cua acumulada dreta, i el fet que no hi hagi cap
abilitats més extremes (a les dues cues).

Cartagena 0 i desviació estàndard 1.5

Centimetres 0 1.5 30000 4 > resu.txt

N: 30000
 Mínim: -5.5962
 Màxim: 6.9732
 Mitjana: 0.0079
 Desviació estàndard: 1.5107

Frec. observada (%)	Frec. esperada (%)	$(\text{Obs-Esp})^2/\text{Esp}$
0.0033	0.0133	0.0075
0.1433	0.1289	0.0016
0.8200	0.8692	0.0028
3.8533	3.8329	0.0001
11.0833	11.0652	0.0000
21.1033	20.9255	0.0015
25.8833	25.9335	0.0001
20.8433	21.0661	0.0024
11.1733	11.2145	0.0002
3.9933	3.9108	0.0017
0.9867	0.8928	0.0099
0.1133	0.1472	0.0078
	Suma=	0.0355
	X2.90=	14.6837 normal
	X2.95=	16.9190 normal
	X2.99=	21.6660 normal

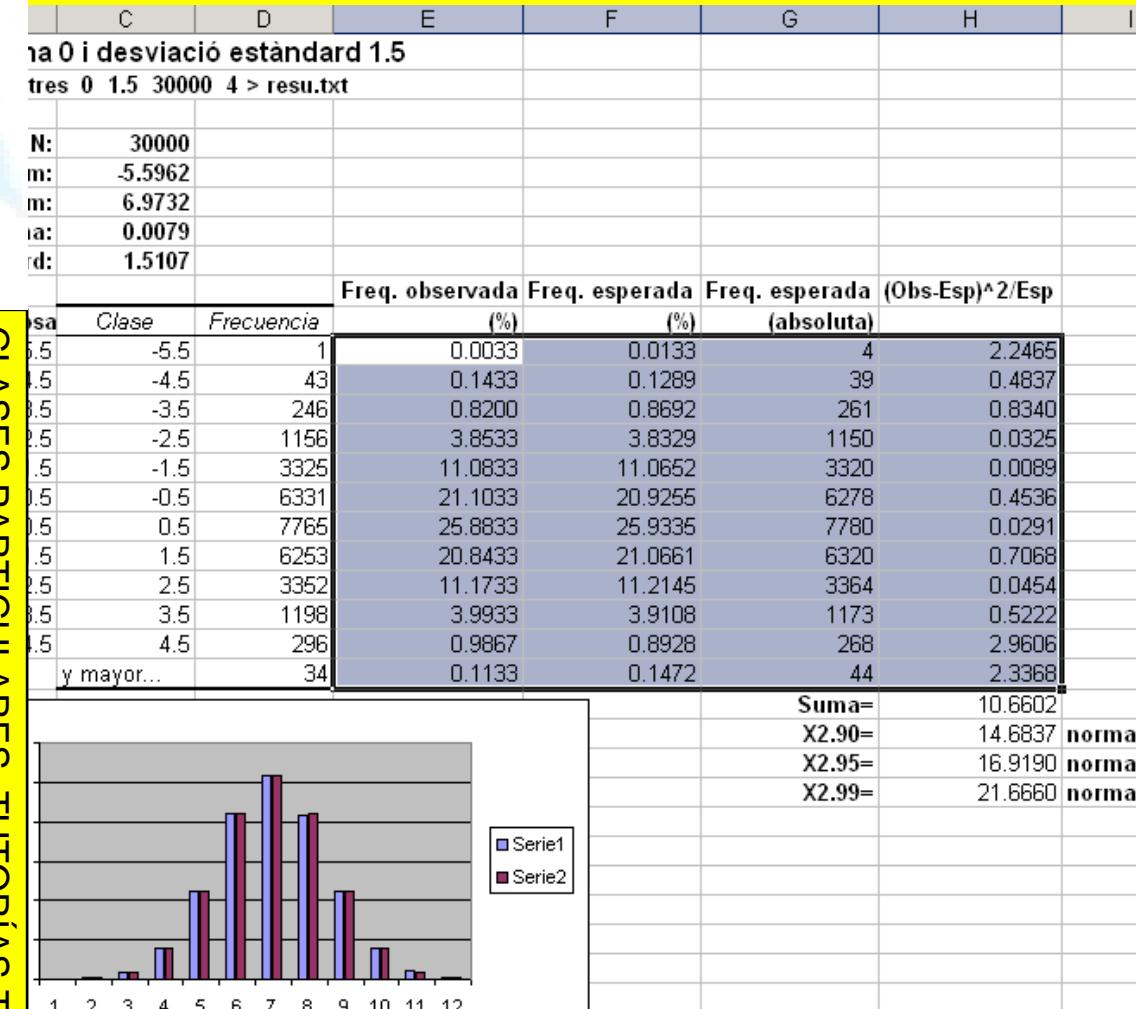


CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70
 ...

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
 CALL OR WHATSAPP: 689 45 44 70

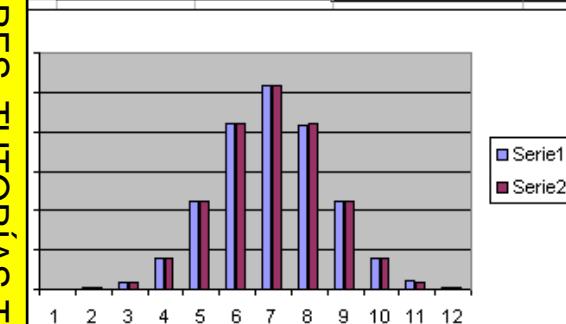
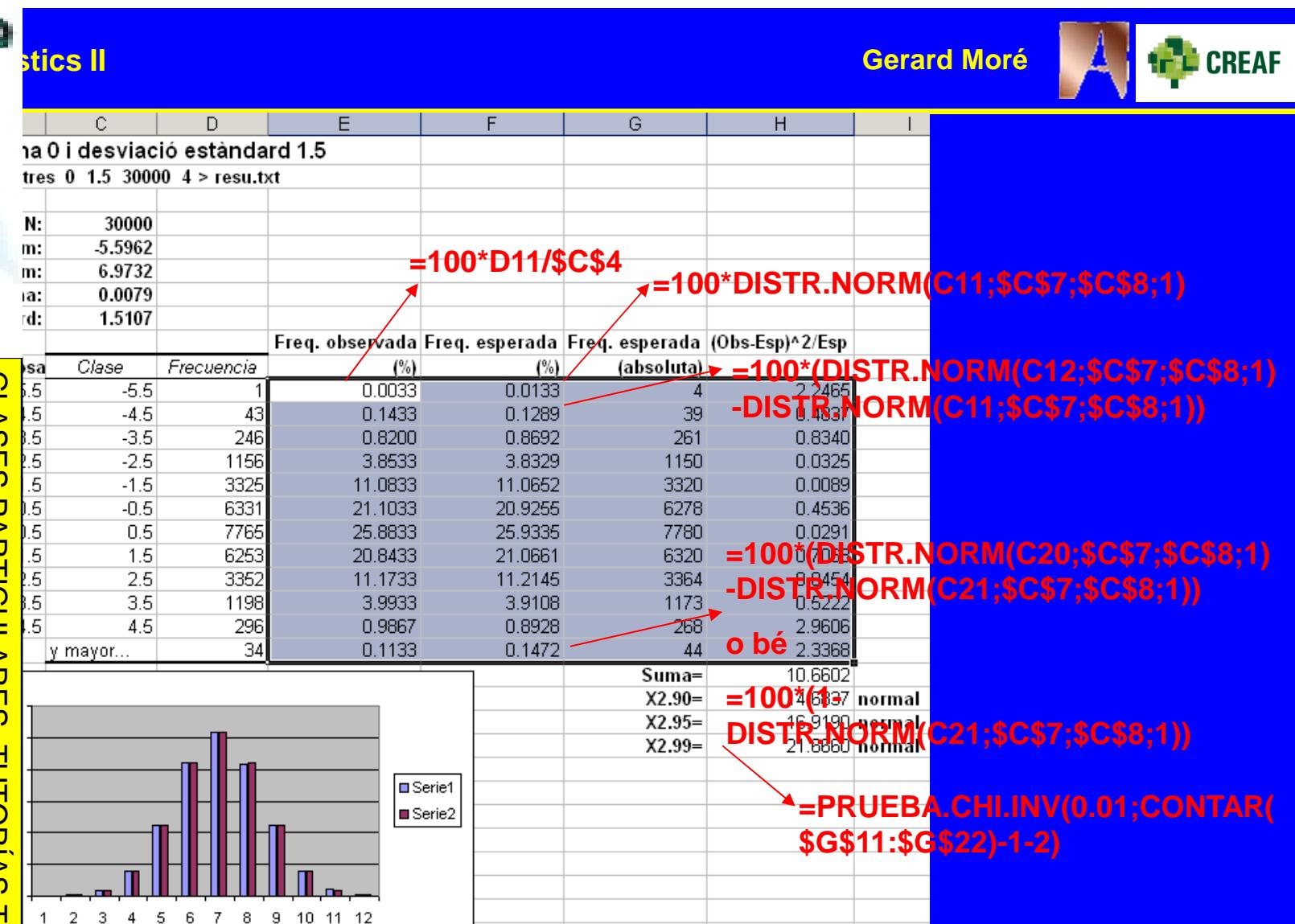
CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70



ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70



Not que sempre s'accepta normalitat. Tanmateix (EXERCICI) proveu què passa amb observarem que si exigim un 5% o un 1% de probabilitat de rebutjar H_0 i que fos certa, no que no és normal. En canvi, podem dir que no és normal acceptant un 10% de risc. A aquest artefacte hi contribueixen les dues freqüències "tallades" a $\frac{1}{2}$ als dos extrems (que no s'ajusten als cues de la normal) i que fins i tot es comparen amb la cua acumulada dreta, i el seu percentil cap dada en les probabilitats més extremes (a les dues cues).

i 1: A partir del Mapa d'Usos del Sòl de Catalunya 1998, disponible al web del Departament de Medi ambient de la Generalitat i del Model Digital d'Elevacions de Catalunya "cata200.img", extret del web de l'Institut Cartogràfic de Catalunya, esbrineu si es pot afirmar que la categoria "Bosquines i prats" segueix una distribució normal a un nivell de significació 0.10, 0.05 i

...

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

Per a obtenir Cata200.img a partir del web de l'ICC o a la ruta <http://www.ica.es/roma\MaterialsMaster\MEII\GerardMore\cata200.IMG> tregeu les altituds a les quals se situa la categoria Bosquines i prats mitjançant una malla de 1000 m x 1000 m.



correlació en imatges de teledetecció

imitat espectral de les bandes de Teledetecció, i la seva
icitat en el cas de bandes que abasten un rang ampli
uds d'ona fa que sovint la correlació entre bandes sigui

rograma bivariant és una excel·lent forma d'apreciar
l'existència d'aquest fenomen.

Invoqueu el mòdul REGRESS d'Idrisi. A partir de la imatge
sat-7 ETM+ de 13-06-2002, compareu la informació de les bandes 3 i
5, 4 i 5, 5 i 6 i 4 i 6. Comenteu els resultats.

Trobar el fitxer a \joanma\materialsMASTER\MEII\XavierPons

rograma bivariant també permet mostrar la presència de
ups naturals (classes) a la imatge. Identifiqueu quins
els histogrames anteriors.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

i 2: Tradicionalment es pensa que l'ús d'imatges de dates sobre una mateixa zona geogràfica aporta més informació útil per a la classificació. Tanmateix, ens podem interesar fins a quin punt la fenologia dels conreus i de la vegetació, la presència de neu i núvols pot aportar més informació que les diferències espectrals. A partir de les imatges del 13-05-2002 i 13-06-2002, compareu la informació de les dates 3, 4, 5 i 6 entre dates. Comenteu els resultats.

a \\joanma\materialsMASTER\MEII\XavierPons

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

de Components Principals (*Pearson 1901*)

Aquestes anàlisis pot ser útil reduir el nombre de variables independents i treballar amb un subconjunt d'elles però amb similar poder informatiu. Això permet, per exemple, dur a terme processos de classificació d'imatges per a la detecció en base a un nombre menor de bandes i, per tant, en menys temps de càlcul.

Ja hem vist que, particularment en les imatges de satèl·lit, la detecció hi ha una alta redundància entre bandes. Per exemple, aquest fet és comú a molts estudis en què s'utilitzen diverses dades en forma multivariant: són ben freqüents les correlacions entre altitud topogràfica i temperatura de l'aire (positiva), o entre superfície plana i pendent topogràfic (negativa).

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

...

de components principals (ACP o PCA) permeten nou conjunt de variables (noves bandes o canals en la Teledetecció) en nombre igual al del conjunt original però amb les següents característiques:

Una nova variable (component principal, CP) s'obté per combinació lineal de totes les p variables originals:

$$C_1 = a_{1,1}X_1 + a_{1,2}X_2 + \dots + a_{1,p}X_p$$

$$C_2 = a_{2,1}X_1 + a_{2,2}X_2 + \dots + a_{2,p}X_p$$

$$\vdots$$

$$C_p = a_{p,1}X_1 + a_{p,2}X_2 + \dots + a_{p,p}X_p$$

Cada nova variable està descorrelacionada amb l'anterior.

Cada nova variable explica una quantitat menor (o igual) de la variancia original total.

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

variables en joc estan molt poc o gens correlacionades, serveix de res.

Finalment no s'obtenen totes les CP (aquesta és la gràcia).

La suma de les variàncies de totes les CP és igual a la suma de variàncies de totes les variables originals (X).

Per d'evitar que alguna de les variables tingui un pes exagerat en l'anàlisi és habitual normalitzar (estandarditzar) les variables (mitjana 0 i desviació estàndard 1). Això permet efectuar l'anàlisi en base a la matriu de correlacions en lloc d'en base a la matriu de variàncies/covariàncies.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

Normalització (estandardització)

(https://en.wikipedia.org/wiki/Standard_score). El llibre d'Spiegel (1991, p. 160), defineix el **standard score** per unitat estàndard, però jo prefereixo valor estàndard que permet usar score com a valor en tots els casos

In statistics, a **standard score** (also called **z-score** or **normal score**) is a dimensionless quantity derived by subtracting the population mean from an individual (raw) score and then dividing the difference by the population standard deviation.

$$z = \frac{x - \mu}{\sigma}$$

The z-score reveals how many units of the standard deviation a case is above or below the mean. The z score allows us to compare the different normal distributions, something done frequently in

This conversion process is called **standardizing**. The quantity z measures the distance between the raw score and the population mean in units of the standard deviation. z is negative when the raw score is below the mean, positive when above.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

procés es calculen els vectors a_1, a_2, \dots, a_p , que reben el nom de vectors propis (*eigenvectors*) i contenen els coeficients del CP.

es calculen uns valors $\lambda_1, \lambda_2, \dots, \lambda_p$, que reben el nom de valors propis (*eigenvalues*) i són la variància de cada CP.

també és interessant perquè dóna idea de quina variabilitat intrínseca de les nostres dades (quantes variacions hi ha).

ves variables perdren les seves unitats i sentit físic. Per exemple, en el cas de bandes de Teledetecció, les bandes ja no tenen el sentit radiomètric que tenien.



eu una ACP de la imatge Landsat-7 ETM+ de 13-
i obtingueu un nombre de CP igual al nombre de
originals.

Invoqueu el mòdul PCA d'Idrisi.

de quina nova CP sembla, visualment, que no hi
ació d'informació rellevant? Com lliga això amb els
propis?

s el paper dels elements singulars però poc
tats a la imatge? Queden preservats?

oveu la descorrelació que s'ha aconseguit en les noves

Useu el mòdul REGRESS d'Idrisi.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

- i 3: Comproveu, en la mateixa imatge, si el fet de fer per un factor 3 una banda mentre dividim entre tres canvia substancialment l'ACP en mode standarditzat i en mode no estandarditzat.
- i 4: Obtingueu de nou les bandes originals a partir d'ACP però obviant les darreres components (menys ives). Arecioeu alguna millora?