# ANÁLISIS ESTADÍSTICO

Estadística descriptivos: Tablas, gráficos, estadísticos descriptivos

Jorge Fallas jfallas56@gmail,com

2010

## Describiendo el set de datos

#### Conocer contexto de los datos

- Variable, nivel de medición
- ¿Qué se quiere resaltar en el set de datos?
- ¿Cuál es la población de referencia?
- Exploración inicial-síntesis
- Tablas y distribuciones de frecuencia
- Análisis gráfico
  - Box-Whisher
  - Gráfico de errores (S, IC)
  - Barras, Histogramas, Polígonos de frecuencia, Ojiva
  - Diagramas probabilidad normal
  - Diagramas de pastel
- Estadísticos descriptivos
  - Tendencia central, variabilidad, forma
  - Datos sin agrupar y agrupados
  - Comparación de series estadísticas

## Tablas y distribuciones de frecuencia

- Clasificar variable

   cuantitativa ó cualitativa
- Variable cuantitativa
  - Definir ancho de clase
  - Número de clases
  - Punto medio de clase
  - Límites reales (variable continua)
- Variable cualitativa
  - Clases definidas por atributos (ej. especie, familia, región, tipo de cobertura, etc.)

## Tablas e histogramas en Excel

1. Seleccione Herramientas, análisis de datos e Histograma

2. Seleccione las celdas que contienen los datos que desea analizar (Rango de entrada)

3. Seleccione las celdas que contienen las clases (Rango de Clases)

4. Rótulos: Seleccione esta casilla si su rango de entrada y su rango de clases incluye la etiqueta o nombre de la columna (Ej. d(cm) ó IC)
5. Porcentaje acumulado: crea frecuencia acumulada

1	A	В	C	D		A	B	C	D	F	F	G	H		I	K	
1	d(dm)	IC	IC	IC IC	1	d(dm)		IC	IC		1						
2	16.7	15	20	14	2	16.7	15	20	14	14	Histogra	ama	_		_	?>	5
3	28.4	20	30	18	3	28.4	20	30	18	16	Entrad	a				Acostas	
4	19.5	25	40	22	4	19.5	25	40	22	18	Rango	de <u>e</u> ntrada:		\$A\$2:\$A\$100		Aceptar	
5	19.9	30		26	5	19.9	30		26	20	Rapgo	de clases:		4842.4846	3	Cancelar	
6	17.6	35		30	6	17.6	35		30	22				14042.4040		0 mida	
7	16			34	7	16			34	24	. 🔽 Ró	tulos				Aygua	
8	18.2				8	18.2				26	Oneine	مرابع مار م					
9	15.6				9	15.6				28		ies de salida			=1		
10	14.8				10	14.8				301	. CRa	ngo de <u>s</u> alida:		L			
11	16.1				11	16.1				32	. 🗭 En	una <u>h</u> oja nuev	/a:	1			
12	31.7				12	31.7				34	C En	un libro nueva	6				
13	10.9	•••••			10	10.9				00	E Par	reto (Histograf	na ordena	do)			
14	14.4				14	14.4				JU		rcentaie acum	ilado			•	
15	15.7				16	24.7						ear gráfico				2	
16	24.7		Análisis de	datos	17	21.9											
17	21.9		Eunciones (	para análisis	140	~~~								·			
18	6.9		Análisis de	e varianza de un fac	ctor	~			12		_						
19	24.2		Análisis de	e varianza de dos fa	actore	es con varias	muestras po	or grupo	1	Cancelar							
20	37.9		Análisis de	e varianza de dos fa	actore	es con una so	ola muestra p	or grupo									
21	22.1		Covarianz	e de correlacion						Ayuda							
22	12		Estadística	a descriptiva					-								
23	14.8		Suavizació	in exponencial	2003272												
24	18.1		Prueba F p	para varianzas de d	tos m	uestras											
25	18		Histogram	a						1							
26	9.1		35 310							(c)							

## Resultado

- 1. Clases con intervalos de 5cm
- 2. Frecuencia absoluta
- 3. Frecuencia acumulada *menor que* (límite superior)
- 4. Histograma y línea de frecuencia acumulada



A partir de esta tabla usted puede calcular frecuencia relativas y frecuencia acumulada mayor que

# **Tendencia** central

- Media aritmética (variables cuantitativas)
  - Suma de variable entre número de observaciones
  - Media geométrica, media armónica
  - Media simple y ponderada
- <u>Mediana</u> (variables cuantitativas) distribución asimétrica o casi simétrica)
  - Observación que indica el 50% de las observación (datos ordenados)
- *Moda* (variables cualitativas)

Valor más frecuente del set de datos



### Variable normalizada



### Distribución de dos variables



#### Comparación gráfica de la distribución de dos muestras

# Variabilidad

- Describe la dispersión del set de datos
- Variables cuantitativas
  - Varianza y desviación estándar
  - Coeficiente de variación
  - Rango: valor máximo valor mínimo
  - Desviación semiintercuartil (Q3-Q1)/2

### **XLStat**

ptions for XLStatistics					
Entering Data					
If you have a range of cells highlighted in a workbook when an XLStatis	tics O Don't automatically enter data				
analysis workbook is selected from the XLStatistics menu, you can choo automatically enter it into the workbook's Data area.	Automatically enter data				
Protection of workbooks and worksheets					
The XLStatistics workbooks are templates that can be corrupted by cert accidental changes to the structure of the workbook (like renaming shee to cells (those with blue text). Unfortunately protecting worksheets also s unprotect individual sheets by using Tools   Protection   Unprotect sheet <b>Copen XLStatistics workbooks with Protection of workbook</b>	ain operations. Protecting the workbooks stops ets); protecting worksheets allows only safe changes tops you from formatting - if necessary, you can - no password is set. <b>.ks turned on</b> <b>ets turned on</b>				
Recording					
You can use Record on the XLStatistics menu to save results in	🔀 Use Ctrl-r as shortcut key for recording				
workbooks with linked data.	Show Record sheet after recording				
Hide Launchpad					
<b>©</b> Open XLStatistics workbooks in <u>R</u> ead-Only Mode	Save current options				
Zoom sheets to fit current windows on opening					
<u></u> Doin succes to ite conclut windows on opening					

Para carga el macro haga un doble clic sobre el archivo XLSTAT.xls. XLSTAT es un conjunto de funciones estadísticas y matemáticas que le permiten visualizar y analizar sus datos. Utiliza como interface de entrada y salida las hojas de cálculo Excel.

## **XLStatistics**

XL	XLStatistics - Excel Workbooks for Statistical Data Analysis © © Rodney Carr 1997-2003									
	Data Analysis Workbooks									
	1	1 Numerical	1Num	1 Categorical	1Cat					
iables	2	1 Numerical 1 Categorical	1Num1Cat	2 Categorical	2Cat	2 Numerical	2Num			
of Var	3	1 Numerical 2 Categorical	1Num2Cat	2 Numerical 1 Categorical	2Num1Cat					
mber (	n	1 Numerical n Categorical	1NumnCat	n Categorical	nCat	n Numerical	nNum			
Ň		n Numerical 1 Categorical	nNum1Cat							
	Other Workbooks									
Probability Functions			PDF	Transform	Transfrm	Options	Help			
Sample Selection			SampSel	Populate	Populate	Hide Launchp	bad			
Qu	alit	y Control	Control			Zoom sheets	to fit window			

### Menú general

## **XLStatistics**

G

#### Microsoft Excel - Book3

	<u>File E</u> di	it <u>V</u> ie	ew	Inser	t F	<u>o</u> rmat	<u>T</u> ools	<u>D</u> ata	<u>W</u> in	dow	Help	XL <u>S</u> t	atistics
2		Pas	steS	<u>a</u>	督	f.	Arial	-	10	• E	8 <u>A</u>		1Num
	A1		-		=	-							1Cat
1 2 3 4 5 6 7 8 9				B		C		D			E	:  	1Num1Cat 2Cat 2Num 1Num2Cat 2Num1Cat 1NumnCat nCat
11													nNum1Cat
12 13 14 15												:	PDF 5ampSel Transfrm
16													Control
18													Record
20 21 22 22													Options Help Quit

#### Menú principal del programa XLStatistics:

1 Num : análisis de una variable numérica

1 Cat : análisis de una variable cualitativa

1Num1Cat : análisis de una variable numérica y una variable cualitativa

#### Una variable numérica (datos de una muestra simple) 1Num

Numerical and graphical summaries

- Sample means, standard deviations...
- Frequency charts (many variations)
- Box-and-whisker plots

#### Tests, etc

- t-test for μ
- Confidence intervals
- Sign test and confidence interval for the media
- Chi-square test for variance
- Residuals plots
- Power analysis/sample size selection
- Test for normality
- Other graphs
  - Bar chart with error bars
  - Cumulative frequency plots of the data or log of the data on a probability scale.

#### Síntesis numéricas y gráficas

- Medias de la muestra, desviaciones estándar...
- Histogramas
- Diagramas (Box-andwhisker plots)

#### Pruebas, etc..

- t-prueba para μ
- Intervalos de confianza
- Prueba del signo y el intervalo de confianza para la mediana
- Prueba de Chi-2 para la varianza
- Gráficos de residuales
- Análisis de potencia y talla de muestra
- Prueba de normalidad

#### Otros gráficos

- Diagrama a barras con barras de error
- Gráficos de frecuencia acumulada o de las probabilidades

Una variable Cualitativa 1Cat	Numerical and graphical summaries • Sample proportions tables, bar charts, etc Tests • Large- and small- sample tests for proportions • Confidence intervals for proportions • Power analysis/sample size selection • Goodness-of-fit test • Runs test Other graphs • Pie chart • Frequency chart with error bars	<ul> <li>Sintesis numéricas y gráficas <ul> <li>Tablas de proporciones, diagramas de barras, etc</li> </ul> </li> <li>Pruebas <ul> <li>Pruebas de proporciones sobre grande y sobre pequeña muestra</li> <li>Intervalos de confianza para proporciones</li> <li>Análisis de potencia y talla de muestra</li> <li>Prueba de adecuación</li> <li>Prueba de Runs</li> </ul> </li> <li>Otros gráficos <ul> <li>Gráfico circular (de quesos)</li> <li>Histogramas con barras de error</li> </ul> </li> </ul>
-------------------------------------	---	---

#### Una variable numérica, una variable cualitativa 1Num1Cat

Numerical and graphical summaries

- Sample means, etc
- Frequency charts
- Box-and-whisker plots
- Plots of means with error bars

#### Tests

- Oneway Analysis of Variance (Random and Fixed effects)
- Test for intercept term
- Kruskal-Wallis test
- Residuals plots
- Hartley's test
- Two-sample t-tests and confidence intervals for the difference between two means
- Mann-Whitney test
- F-test
- Power analysis/sample size selection (2 level only)

Other graphs

• Means plots with error bars or bands of various

#### Síntesis numéricas y gráficas

- Medias de la muestra, etc..
- Diagramas de frecuencia
- Gráficos de caja (Box-andwhisker plots)
- Gráficos de lss medias con barras de error

#### Pruebas

- Análisis de la varianza simples (efectos aleatorios y fijos)
- Determinación del corte
- Prueba de Kruskal-Wallis
- Gráficos de los residuales
- Prueba de Hartley
- T-pruebas sobre dos muestras e intervalos de confianza para la diferencia entre medias
- Prueba de Mann-Whitney
- F-prueba
- Análisis de la potencia y la talla de muestra (nivel 2 solamente)

#### Otros gráficos

- Trazado de las medias con barras de error
- Histogramas

## XLStat: Análisis de 1 variable

1Num
Analysis of Data for a Single Numerical Variable
Data goes in the Data area on the Data & Description sheet. Double-clicking on the cell marked Data
(at the top of the Data area) brings up help.
Sheets
Data & Description: Data and some basic descriptions/summaries.
Summaries: Summaries appropriate for this type of data.
Tests: Statistical tests, etc, appropriate for this type of data.
Extra Tools: A workbook for analysing grouped data.
Cells that you may alter are blue - others should not be changed. Do not Cut, Move or Delete cells.
Do not change the name of the workbook.
Charts may be edited (as usual) if necessary (to set colors, place legends, etc) but do not delete
series or legend entries otherwise the updating routines may fail. When copying charts to other
applications use Paste Special   Picture.
Help is obtained by double-clicking on any red cell. Read XLStats.doc for more information and
examples.

Rodney Carr, 1997-2006



### XLStat: Análisis de 1 variable

Information	Data & Description

Inserte los datos en

la columna "DATA".

**Recuerde que solo** 

Data

d (cm)

16.728.4

#### Diagrama de barras e histograma



Límite inferior de la primer Límite superior de la última - Graficar Límite superior de la clase superior ó los puntos medios de las clases Si grafica variables continuas debe marcar la casilla de Histograma

(Histogram)

### Data & Description / Descripción de datos



En esta sección XLStats muestra:

#### 1. Estadísticos descriptivos

2. Barra/Histograma Usted puede modificar cualquiera de los números en color azul. Ejemplo:

- 1. Número de clases
- 2. Límite inferior
- 3. Límite superior

3. Grafico de variable según orden



### Resumen



#### Frequency Chart

Box and Whisker Plot



En esta sección el programa le muestra:

1. Grafico de media y una medida de variabilidad

2. Barra/Histograma/Línea Usted puede modificar el número de clases, valores mínimo y máximo y decimales.

Etiquetas: punto medio de la clase ó clases

<u>Estilo</u>: barras, histograma, línea

<u>Eje Y</u>: frecuencia, proporción, %

Grafico de Box y Whisker

## Polígono de frecuencia y Ojiva



Usted puede graficar los datos como un polígono de frecuencia seleccionando la casilla "line" en Estilo.

Puede graficar frecuencia absoluta, proporciones y %

Para graficar una ojiva (frecuencia acumulada) selecciona la casilla "cummulative"

#### Box and Whisker Plot

#### Grafico de Box y Whisker



El programa crea una grafica vertical y otra horizontal





Normal Probability Plots 🖻





Grouped Data

⇒

Usted puede modificar los valores en <u>azul</u>.

- 1. Límite inferior de la clase
- 2. Límite superior de la clase
- 3. Frecuencia de la clase



# Diagrama dispersión



El diagrama de dispersión muestra la relación entre la variables X y Y.

Se utiliza para visualizar la correlación o asociación entre dos variables cuantitativas

# Gráfico de error



Variante del gráfico de dispersión. Cada segmento de línea representa el "error" asociado a cada observación. Por ejemplo, se puede expresar el error asociado a la estimación de longitud total de cocodrilos por sitio de muestreo ó el error de estimación de la precipitación en dos o más estaciones pluviométricas

## Pictogramas



Se utiliza un elemento gráfico para representar cantidades

El lector es influenciado por el <u>área</u> del objeto y no por sus dimensiones lineales (altura y longitud)

# Gráfico radar



Objetivo: comparar el comportamiento de tres o más variables en el tiempo.

Las unidades deben ser las mismas (Ej. porcentaje, ha, metros, etc.)

#### Ver ejemplos de gráficos en Excel

### Distribución de frecuencia y desviación estándar de la media.



Distribución de frecuencia esperada para una población de árboles con un diámetro medio de 17.4cm y una varianza de 25cm2. Cualquier parcela con un diámetro promedio superior a 27.5cm ó inferior a 7.5cm se consideraría un evento raro dado la distribución de referencia

## Forma de la distribución

- Describe la forma de la distribución que caracteriza al set de datos
  - Curtosis: grado de achatamiento de la distribución
    - Leptocúrticas
    - Mesocúrticas
    - Platicúrticas
  - Asimetría: grado de desplazamiento del centro de gravedad del set de datos
    - Simétricas
    - Asimétricas
- Ver histogramas para evaluar asimetría y curtosis visualmente
- Ver coeficiente de curtosis y de asimetría para su evaluación numérica

## Percentiles y cuantiles

- Percentiles: se divide la serie estadística en centésimos (0-100%).
  - Ej. Percentil 40 es aquella observación por debajo de la cual se encuentra el 40% de las observaciones. Uso de grafica de OJIVA: frecuencia acumulada en %

Mediana percentil 50

 <u>Cuantiles</u>: división de la serie estadística relativa a cualquier partición (Ej. terciles, cuartiles, quintiles, deciles)

## Cuartiles



#### **RESULTADO**

Ê	CUARTIL	- X V	= =CUART	TIL(A2:A10	)0;1)
Ũ.,	A	В	C	D	E
1	d(cm)	valor	cuartil		
2	16.7	=CUARTIL	(A2:A100;1)		
3	28.4	17.5	2		
4	19.5	21.3	3		1
5	19.9	37.9	4	Ş	
6	17.6				
7	16				

En la casilla respectiva reemplace =CUARTIL(A2:A100,1) cuartil 1 =CUARTIL(A2:A100,2) cartil 2 =CUARTIL(A2:A100,3) cuartil 3 =CUARTIL(A2:A100,4) cuartil 4

### PERCENTILES

P	ERCENTIL	▼ X ✓ = =PERCENTIL(A2:A100;0.1)							
Ü., 1	A	В	С	D	E				
1	d(cm)	valor cm	percentil	Manaka kata kata kata k					
2	16.7	=PERCEN	TIL(A2:A10	0;0.1)					
3	28.4	12.2	2						
4	19.5	14.48	3						
5	19.9	15.76	4	2) 5)					
6	17.6	17.5	5						
7	16	18.18	6						
8	18.2	19.86	7	619191919191919191919 2					
9	15.6	22.04	8	onenenenenenen S Marananananananan					
10	14.8	24.12	9						
11	16.1	37.9	10						
12	31.7								

En la casilla respectiva digite =PERCENTIL(A2:A100,0.1) Percentil 10 =PERCENTI (A2:A100,0.2) Percentil 20 = PERCENTIL(A2:A100,0.3) Percentil 30 = PERCENTIL(A2:A100,0.4) Percentil 40 = PERCENTIL(A2:A100,0.5) Percentil 50 = PERCENTIL(A2:A100,0.6) Percentil 60 = PERCENTIL(A2:A100,0.7) Percentil 70 = PERCENTIL(A2:A100,0.8) Percentil 80 = PERCENTIL(A2:A100,0.9) Percentil 90 = PERCENTIL(A2:A100,1) Percentil 100

#### 1. Resumiendo los Datos

#### Estadística Descriptiva

Número de observaciones, Max, Media, Min, Varianza, Rango, Desviación Estándar, Primer Cuartíl, Coef. de Variación, Coef. de Var. Cuartíl, Asimetría (Oblicuidad), Tercer Cuartíl, Kurtosis, Rango Intercuartíl (RIC = Q3 -Q1), CVC = (Q3-Q1) / (Q3 + Q1). Error Est. Media, Desviación Cuartíl, Mediana, Media Abs. de Desv. Histograma : Anchura de la clase, Número de clases, Puntos medios de la clase, frecuencias Determinación de los valores extremos (Outliers) Función de Distribución Empírica Histogramas Índice de Estacionalidad (series de tiempo)

Las Tres Medias : Media Aritmética, Media Geométrica, Media Armónica

#### 2. Cálculo de Probabilidad

<u>Comparando Dos Variables Aleatorias</u> <u>Distribución Multinomial</u> <u>Valores P para la Distribución Populares</u>

# Pruebas de normalidad

### • W de Shapiro-Wilk

- Considerada una de las pruebas mas fidedignas o imparciales
- D de Kolmogorov-Smirnov
- Prueba de bondad de ajuste de una distribución
- Lillifors
- Versión modificada de la D de Kolmogorov-Smirnov
- Ho: los datos provienen de una distribución normal
- Ha: los datos no provienen de una distribución normal
- Gráfico de probabilidad normal