



VNIVERSITAT ID VALÈNCIA

MASTER DE INGENIERÍA BIOMÉDICA.

Guía para la realización de los trabajos (II)

Representaciones gráficas; estadísticos y contraste de hipótesis

Profesores: Emilio Soria y Antonio José Serrano, Dpto Ingeniería Electrónica, ETSE

VNIVERSITAT ID VALÈNCIA

Representaciones gráficas

En primer lugar hay que cargar el conjunto de datos (la manera más sencilla es siguiendo la secuencia File-Import Data). También lo podemos hacer desde el teclado; se cargará un fichero de ejemplo y se mostrarán las diferentes instrucciones sobre los datos cargados. Ejecuta la siguiente instrucción `load iris.dat` (comprueba qué variables has cargado!!). Deberías acceder a la ayuda de la instrucción y ver las diferentes posibilidades que te ofrece la instrucción y, lo más importante, saber qué estás viendo.

Gráfico	Instrucción	Ejemplo
Histograma	<code>hist(x)</code>	<code>hist(iris(:,1))</code>
Diagrama de caja	<code>boxplot(x)</code>	<code>boxplot(iris(:,1:3))</code>
Gráfico de cuantiles	<code>qqplot(x)</code>	<code>qqplot(iris(:,1))</code>
Gráfico de probabilidad	<code>probplot(x)</code>	<code>probplot(randn(1,1000))</code>

Estadísticos.

Seguimos con el ejemplo del fichero iris (cárgalo antes de ejecutar la columna de ejemplo); recuerda que, si el argumento es una matriz, estas instrucciones devuelven un vector correspondiente al estadístico correspondiente calculado para cada columna.

Estadístico	Instrucción	Ejemplo
Valor medio	<code>mean(x)</code>	<code>mean(iris(:,1:3))</code>
Desviación estándar/varianza	<code>std(x)/var(x)</code>	<code>std(iris(:,1:3))/var(iris(:,1:3))</code>
Sesgo	<code>skewness(x)</code>	<code>skewness(iris(:,1:3))</code>
Kurtosis	<code>kurtosis(x)</code>	<code>kurtosis(iris(:,1:3))</code>
Mediana	<code>median(x)</code>	<code>median(iris(:,1:3))</code>
Percentiles	<code>prctile(x,p)</code> ; p indica el percentil a calcular	<code>prctile(iris(:,1),25)</code>
Rango intercuartil	<code>iqr(x)</code>	<code>iqr(iris(:,1))</code>
Rango	<code>range(x)</code>	<code>range(iris(:,2))</code>

Variables discretas

En el caso de tener variables discretas las anteriores representaciones gráficas y estadísticos se complementan con las siguientes instrucciones

tabulate(x); devuelve una matriz con tres columnas; la primera corresponde a los casos que se tienen; la segunda informa sobre el número de veces que aparece cada caso y la tercera da el porcentaje de aparición. Ejecuta la siguiente instrucción `tabulate(iris(:,5))` e intenta comprender la información que aparece en pantalla.

crosstab(x,y); sirve para construir de forma automática las tablas de contingencia que relacionan dos variables discretas (por ejemplo sexo y tener (o no) una determinada enfermedad); cuando se llegue al contraste de hipótesis se volverá a repasar esta instrucción. Ejecuta la siguientes instrucciones e intenta comprender lo que se obtiene; `sexo=round(rand(1,10)),enfermo=round(rand(1,10)), crosstab(sexo,enfermo)`

En cuanto a representaciones gráficas, el diagrama de sectores es el más utilizado; la instrucción en Matlab es **pie(X)**. Generalmente se usa conjuntamente con la instrucción **tabulate**. Comprueba lo comentado con la siguiente secuencia de instrucciones:

```
x=tabulate(iris(:,5)),pie(x)
```

Varias variables (continuas)

Si se quieren relacionar varias variables continuas, el primer parámetro para analizar es el **coeficiente de correlación** que determina si existe (o no) una relación lineal entre las variables. En Matlab se determina usando la instrucción **corrcoef(X)**. **RECUERDA SIEMPRE MIRAR LA AYUDA!!!!**. Ejecuta la siguiente instrucción **corrcoef(iris(:,1:3))**, **¿existe relación lineal entre las variables?**; **¿existe alguna relación entre las variables?**. Ejecuta ahora las siguientes instrucciones **x=randn(100,5); y=corrcoef(x); imagesc(y); colorbar** (como siempre...mira la ayuda!!!!), intenta interpretar lo obtenido.

Siempre que se puede hay que encontrar una representación gráfica de las variables de entrada ya que una buena gráfica proporciona más información que cualquier estadístico.

Una instrucción que ofrece información visual sobre las variables de forma individual y conjunto es **plotmatrix(x)**. Ejecuta la instrucción **plotmatrix(iris(:,1:3))** y observa lo que se obtiene (tienes que interpretar el resultado como si fuera una matriz).

Contraste de hipótesis

Para comprobar las diferentes instrucciones vamos a generar un conjunto de vectores que serán los siguientes: a) $x_1 = \text{randn}(1, 1000)$; b) $x_2 = \text{rand}(1, 1000)$; c) $x_3 = 2 * x_1$. Hay que mirar la ayuda pues en la tabla se muestra *el uso básico* de las diferentes instrucciones

Uso	Instrucción	Ejemplo
Comprobar la normalidad (valor medio cero, varianza=1)	$[h,p] = \text{kstest}(x)$	$[h,p] = \text{kstest}(x_1)$ $[h,p] = \text{kstest}(x_2)$
Comprobar la normalidad (cualquier media y varianza)	$[h,p] = \text{jbtest}(x)$	$[h,p] = \text{jbtest}(x_1)$ $[h,p] = \text{jbtest}(x_2)$
Hipótesis nula=los dos vectores proceden de la misma distribución	$[h,p] = \text{kstest2}(x,y)$	$[h,p] = \text{kstest2}(x_1,x_2)$ $[h,p] = \text{kstest2}(x_1,x_2)$
Hipótesis nula=los dos vectores tienen la misma distribución e idénticas medianas	$[p,h] = \text{ranksum}(x,y)$ CUIDADO AQUÍ VA AL REVÉS	$[p,h] = \text{ranksum}(x_1,x_2)$ $[p,h] = \text{ranksum}(x_1,x_3)$ $[p,h] = \text{ranksum}(x_1,(x_3)+3)$
Hipótesis nula=los dos vectores siguen una distribución normal con las mismas medias	$[h,p] = \text{ttest2}(x,y)$	$[h,p] = \text{ttest2}(x_1,x_2)$ $[h,p] = \text{ttest2}(x_1,x_2)$

Matlab dispone de una instrucción para determinar la independencia en tablas de contingencia; esta instrucción es **crosstab(x,y)**

Ejecuta las siguientes instrucciones: $x_1 = \text{round}(\text{rand}(1, 100))$; $x_2 = \text{round}(\text{rand}(1, 100))$; $[\text{tabla}, \text{chi2}, p] = \text{crosstab}(x_1, x_2)$; ¿qué ocurre si el argumento es (x_1, x_1) ?

Profesores: Emilio Soria y Antonio José Serrano, Dpto Ingeniería Electrónica, ETSE

UNIVERSITAT DE VALÈNCIA

Ahora a practicar....de verdad!!!.....

En la página web de la asignatura tienes los siguientes ficheros

Nilo.txt	Datos de flujo del río Nilo durante el periodo comprendido en los años 1871-1970
cemento.txt	El fichero contiene la cantidad de calor necesaria (última columna) para crear 13 tipos de cementos diferentes; las cuatro primeras columnas son las proporciones de los elementos usados en su creación.
orange.txt	Este fichero tiene variables que están relacionadas con el crecimiento de los naranjos; se tienen tres columnas que son: a) un indicador del árbol; b) edad desde el 31/12/1968 y c) la circunferencia del tronco
urine.txt	Toda la información del fichero aparece en el documento <i>urine.doc</i>
bupa.txt	Toda la información del fichero aparece en el documento <i>bupa.doc</i>

Intenta poner en práctica lo aprendido (cálculo de estadísticos, variables normales?; dependen linealmente?; representación.....

Profesores: Emilio Soria y Antonio José Serrano, Dpto Ingeniería Electrónica, ETSE

UNIVERSITAT DE VALÈNCIA



VNIVERSITAT VALÈNCIA

MASTER DE INGENIERÍA BIOMÉDICA.

Guía para la realización de los trabajos (II)

Representaciones gráficas; estadísticos

Profesores: Emilio Soria y Antonio José Serrano, Dpto Ingeniería Electrónica, ETSE

VNIVERSITAT VALÈNCIA