

Curs pràctic de Maple

Pràctica 14

14 Estadística descriptiva i representació gràfica de dades

Maple disposa del paquet anomenat **stats** que recull un seguit de subpaquets cadascun adequat per a diferents tipus d'anàlisis estadístics de dades. En aquesta pràctica farem un cop d'ull als subpaquets **describe**, **statsplots** i **transform**.

Les dades que es vulguin analitzar estaran guardades en una llista sobre la que aplicarem les diferents funcions del paquet. Si volem estudiar estadísticament els resultats del llançament d'un dau de sis cares cinc cops i els resultats obtinguts són 1, 3, 4, 3, 2 posarem

```
> with(stats):  
> data:=[1,3,4,3, 2];
```

i podrem aplicar a la llista `data` les diferents funcions.

En cas que l'ordre de les dades sigui irrellevant podem incorporar a la llista la funció `Weight(valor,pes)` on el paràmetre `valor` indica un dels valors de la llista i `pes` és el nombre de vegades que el valor apareix a la llista (com si poséssim `valor$pes`).

```
> [1,46, Weight(3,2), 90];
```

Maple ens permet importar un fitxer de dades ja existent mitjançant la comanda `importdata("fitxer",n)` on `fitxer` és l'adreça del fitxer de dades (que serà un fitxer text on les dades estan organitzades per columnes) i `n` és el número de columnes del fitxer. Maple importa el fitxer com una seqüència de n llistes de dades, una per cada columna. En l'exemple següent importem un fitxer on hi ha guardat el temps de vida (en hores) d'una mostra de llums fluorescents que està en el directori `Mis documentos` del disc `c:` del nostre sistema (aquest fitxer el podeu trobar a <http://mat.uab.es/gguasp/PractInt/Lamp/tempssvida.dat> i s'ha de guardar en algun lloc accessible pel sistema operatiu).

Si feu

```
> llums:=importdata("c:/Mis documentos/tempssvida.dat",1);
```

(o el que correspongui pel lloc on heu guardat el fitxer) tindreu en `llums` les dades recollides del que han durat aquests llums.

14.1 Les comandes de **describe**

Habitualment iniciem l'anàlisi de les dades amb un estudi descriptiu. En aquest punt és el subpaquet **describe**, el que ens proporciona les funcions necessàries com són, per exemple, `count` (que compta el nombre de valors que apareixen en la llista de dades), `mean` (que calcula la mitjana aritmètica de les dades introduïdes), `mode` (que dóna com a resultat el valor que apareix més cops dins del conjunt de dades), `variance` (que dóna una mesura de la dispersió dels valors de les dades calculant la mitjana aritmètica dels quadrats de les diferències de cada valor amb la mitjana de les dades) i moltes altres que podreu consultar a l'ajuda de Maple a partir de `help(describe)`;

Exemple 14.1

```
> with(stats):  
> data:=[1,1,3,Weight(2,4)];  
> describe[count](data);  
> describe[mean](data);  
> describe[mode](data);  
> describe[variance](data);
```

Noteu que, de fet, `describe` és una `table` de Maple en la que cada un dels seus elements és una funció diferent que es pot aplicar al conjunt de dades `data`.

Podem agrupar un seguit d'aquestes comandes per formar una llista d'operacions que s'apliquen conjuntament sobre la mateixa llista de dades i així obtenir el conjunt dels valors significatius en un sol resultat.

Exemple 14.2

```
> descriptiva:=[describe[mean],describe[median],describe[variance]];
> descriptiva(data);
```

Exercici 14.1

Appliqueu la comanda `descriptiva` a les dades del temps de vida dels llums fluorescents guardats en `llums`.

Exercici 14.2

Construiu una llista d'operacions que calculi els quartils de les dades de la llista `llums`. Els quartils d'una llista de dades són els tres valors que divideixen el conjunt total de dades, un cop s'han ordenat de menor a major, en quatre parts amb el mateix nombre de dades en cada un d'ells (comptant amb les repeticions que hi pot haver). Busqueu informació sobre `quartile` en l'ajuda de Maple.

14.2 Representació gràfica de les dades

Dins el paquet `stats` l'apartat que conté les funcions per a fer representacions gràfiques és `statplots`

14.2.1 Gràfics boxplot

Les representacions gràfiques resumeixen alguns aspectes de les dades. Un tipus de gràfic adequat per la representació de dades quantitatives és l'anomenat gràfic BoxPlot (diagrama de caixes) del que en teniu un exemple a continuació.

Exemple 14.3

```
> Xdata := [4.535, 4.029, 5.407, 1.605, 5.757, 3.527,
> 7.890,8.159, 6.092, 13.442, 2.845, 5.172, 3.277, 8.810, 3.657, 7.226,3.851,
> 2.162, 2.668, 4.692];
> statplots[boxplot](Xdata);
```

la línia central de la caixa representa la mediana (el valor central de les dades), les línies superior i inferior representen el primer i el tercer quartil i les línies que surten cap dalt i cap baix s'estenen fins a $3/2$ del rang de valors entre quartils.

Noteu que Maple situa el diagrama de caixa sobre la posició zero de l'eix horitzontal. Si volem que el desplaci podem utilitzar l'opció `shift=n`, on `n` és el número d'unitats que volem que es desplaci. A més podem indicar l'amplada de la caixa mitjançant l'opció `width=n`. Si a la comanda `boxplot` introduïu més d'un fitxer de dades

```
> statplots[boxplot](Xdata1,Xdata2);
```

obtdreu els diagrames de caixa corresponents en un sol gràfic. Aquesta opció és útil quan es vol comparar el comportament de més d'un conjunt de dades.

Exercici 14.3

Feu un diagrama de caixa del conjunt de dades del fitxer `tempsvida.dat`

14.2.2 Histogrames

La distribució de les freqüències dels valors observats podem representar-la amb un histograma de freqüències com en l'exemple següent:

Exemple 14.4

```
> statplots[histogram](Xdata, numbars=5, area= 100);  
> statplots[histogram](Xdata, numbars=5, area= count, color=cyan);
```

Observeu que l'opció `numbars=5` fa que es distribueixen les dades en cinc grups, que cada una de les barres té la mateixa amplada i una altura proporcional al nombre de dades que cauen dins el grup corresponent i que el valor numèric d'aquesta altura ens dóna els percentatge de dades en cada grup quan posem l'opció `area=100` i el nombre total de dades quan l'opció és `area=count`. Noteu també que s'hi poden afegir opcions gràfiques com per exemple `color=cyan` que farà que el color del dibuix sigui cyan.

Exercici 14.4

Feu ara un histograma en el que es representin els temps de vida dels fluorescents dividits en 10 grups i de tal forma que l'altura de les barres representi el nombre de mesures en cada un dels intervals.

Exercici 14.5

Importeu el fitxer `data3.dat` que podreu trobar a <http://mat.uab.es/gguasp/PractInt/Lamp/data3.dat>. Aquest fitxer conté tres columnes que corresponen al temps de vida en hores de tres tipus de fluorescents segons el seu cebador: ràpid, de pre-escalfament i instantani. Anomeneu els tres tipus de dades com `rapid`, `escalfament` i `instantani`.

- Calculeu el número de fluorescents de cada tipus estudiats.
- Per cada classe de fluorescent calculeu la mitjana, moda i variància del seu temps de vida.
- Feu un histograma de 10 barres per cada tipus de fluorescent en que les columnes representin la freqüència de casos.
- En un mateix gràfic dibuixeu els corresponents diagrames de caixa i analitzeu el resultat.

14.3 La comanda `rand()`

Maple permet la generació de nombres aleatoris mitjançant la funció `rand()`.

```
> exp1:=rand(a..b);  
> exp2:=rnd(a);
```

La primera línia crea una funció que genera un número aleatori enter entre a i b . La segona línia crea una funció que genera un número aleatori entre 0 i $a - 1$. Comproveu que passa quan les executeu amb valor concrets per a a , b ,

```
> exp1();  
> exp2();
```

Exemple 14.5

Podem simular el llançament d'una moneda fent que el Maple generi números aleatoris enters entre 0 (=cara) i 1 (=creu). Si volem guardar 20 llançaments de la moneda, podem crear la seqüència que els conté.

```
> moneda:=rand(2);  
> sim:=[seq(moneda(),i=1..20)];
```

14.4 Les comandes de **transform**

Finalment analitzarem algunes de les comandes del subpaquet **transform**. En particular veurem com ordenar un conjunt de dades i com comptar el número de vegades que apareix un mateix valor.

La comanda `transform[statsort](data)` serveix per ordenar un conjunt de dades (`data`) en ordre creixent.

Exemple 14.6

```
> ord:=transform[statsort](sim);
```

La comanda `transform[tally](data)` agrupa totes les dades que prenen el mateix valor mitjançant l'expressió `Weight`. D'aquesta manera podem observar el número de vegades que apareix cada valor.

Exemple 14.7

```
> agrup:=transform[tally](sim);
```

Finalment, i un cop hem agrupat les dades, la comanda `transform[frequency](data)` retorna una llista amb la freqüència de cada valor de la llista (número de vegades que es repeteix cada valor).

Exemple 14.8

```
> freq:=transform[frequency](agrup);
```

Exercici 14.6

- i) Simuleu l'experiment que consisteix en llançar un dau 1000 vegades. Calculeu el número de vegades que ha sortit cada cara del dau.
- ii) Fixeu vos que la comanda `histogram` de `statplots` està pensada per agrupar les freqüències d'uns resultats en els que els valors possibles són valors reals però que en el cas d'un experiment en el que els resultats siguin només uns pocs valors (com en el cas del dau o de la moneda) no funcionarà massa bé (proveu-ho). Definiu un procediment que, partint del resultat d'un experiment amb valors discrets (dau, moneda...), doni un gràfic de barres amb les freqüències de cada un dels resultats possibles. Apliqueu-la als resultats de la simulació del llançament del dau. Són més o menys igual d'altres totes les columnes? El dau està trucat?