

Pràctiques Integrades 1er de Matemàtiques

Aproximacions i ajusts de dades

Curs 2003–04

Índex

V	Aproximacions i ajusts de dades	3
16	Aproximació de les arrels d'una equació	3
16.1	Mètode de la bipartició	3
16.2	Mètode de Newton	4
17	Aproximació de funcions	9
17.1	Polinomi interpolador	9
17.2	Spline	10
17.3	Exercicis	10
17.4	Polinomi de Taylor	11
18	Ajust a les dades. Estadística descriptiva	15
18.1	Ajust a les dades	15
18.1.1	Llei de Hooke	15
18.1.2	El mètode dels mínims quadrats	16
18.2	Estadística descriptiva	19
18.2.1	Les comandes de <code>describe</code>	19
18.2.2	Representació gràfica de les dades	20
18.2.3	Gràfics <code>boxplot</code>	20
18.2.4	Histogrames	21
18.2.5	Les comandes de <code>transform</code>	22

Part V

Aproximacions i ajusts de dades

16 Aproximació de les arrels d'una equació

L'objectiu d'aquesta secció és fer un petit anàlisi amb Maple de dos mètodes iteratius que ens permeten trobar aproximacions numèriques de zeros de funcions. Veurem els seus avantatges i desavantatges. Abans però recordeu que a la secció 4 ja havíem analitzat la comanda `fsolve` que ens fa aquesta feina. El que farem avui és construir els nostres propis procediments.

16.1 Mètode de la bipartició

El primer mètode que analitzarem és conegut com el mètode de la bipartició. És una aplicació directa del teorema de Bolzano. Aquest teorema diu que si f és una funció contínua a l'interval $[a, b]$ tal que $f(a)$ i $f(b)$ són no nuls i prenen valors de signe oposat, aleshores existeix un punt dins de l'interval en el qual la funció pren el valor zero. El mètode de la bipartició utilitza aquest fet per a donar intervals cada cop més petits que contenen el zero de la funció.

Anem a implementar aquest mètode amb Maple per a trobar una aproximació numèrica de $\sqrt{2}$ pensada com a zero de la funció $f(x) = x^2 - 2$.

Primer de tot cal esbrinar quin és l'interval convenient per començar. Per això ens ajudarem amb una representació gràfica de la funció.

Exercici 16.1

Mitjançant una representació gràfica de la funció $f(x) = x^2 - 2$ determineu un interval que contingui $\sqrt{2}$.

Noteu que per exemple podem prendre l'interval $[1, 2]$ i comproveu que $f(1)$ i $f(2)$ prenen valors de signe contrari (1 i 2 són cotes inferior i superior respectivament per un zero de la funció). Per tant, podem concloure que $1 \leq \sqrt{2} \leq 2$. Sigui m_1 el punt mig de l'interval, és a dir, $\frac{1+2}{2} = 1.5$.

Exercici 16.2

En quin dels dos intervals està continguda $\sqrt{2}$, $[1, m_1]$ o $[m_1, 2]$?

Ja haureu observat que $f(1)$ i $f(1.5)$ prenen valors de signe contrari. Per tant, l'interval $[1, 1.5]$ conté un zero de la funció. D'aquesta manera hem obtingut un interval més precís que l'inicial.

Noteu que el procés anterior el podeu anar iterant, és a dir, donat aquest nou interval que conté el zero, considereu el punt mig i preneu el subinterval on els extrems prenen valors de signe contrari.

Exercici 16.3

Repetiu el procés anterior tres cops.

Com ja haureu observat aquest mètode recursiu d'aproximació és molt senzill de programar.

Exercici 16.4

Realitzeu un procediment tal que donada una funció f , un número natural n i els extrems a i b d'un interval ens retorni un missatge dient si hi ha una arrel o no en l'interval $[a, b]$ i, si n'hi ha, que iteri el mètode de la bipartició n cops retornant les cotes inferior i superior de la solució de $f(x) = 0$ que s'obtenen.

Exercici 16.5

Modifiqueu el procediment anterior de manera que enlloc del paràmetre n , nombre d'iteracions, tingui com a paràmetre la precisió en decimals que volem obtenir. Una precisió de 10^{-n} vol dir que la diferència entre la cota superior i la inferior és menor que 10^{-n} .

Exercici 16.6

Utilitzeu els procediments anteriors per a calcular una aproximació numèrica amb una precisió de 15 decimals de la solució de $x - \sin(x) - 5 = 0$.

16.2 Mètode de Newton

El mètode de la bipartició abans esmentat és un mètode senzill i fiable però molt lent, és a dir, calen moltes iteracions per a poder obtenir una precisió elevada. El mètode de Newton que tot seguit anem a descriure és molt més ràpid però té altres inconvenients com veurem més endavant. Aquest mètode utilitza la idea de l'aproximació lineal de la funció mitjançant la recta tangent (és a dir, donat un punt inicial considera com a primera aproximació del zero de la funció el zero de la seva recta tangent en aquest punt). Començarem la iteració en un punt x_0 proper al zero de la funció. Considerarem un nou punt que

és el zero de la recta tangent a f en el punt x_0 . Recordeu que la recta tangent a la gràfica de f al punt x_0 és $r(x) = f(x_0) + f'(x_0)(x - x_0)$.

Exercici 16.7

Per a simplificar els càlculs posteriors serà convenient definir una funció `tangent` que tingui com a paràmetres el nom d'una funció (`f`), el punt `a` on es vol calcular la tangent al gràfic de `f` i un valor `x` qualsevol i doni com a resultat el valor corresponent a l'expressió $f(a) + f'(a)(x - a)$.

Per a comprovar que la funció dóna un resultat correcte proveu si les comandes

```
> tangent(cos,0,x);
> tangent(sin,0,Pi);
> tangent(x->1/(1+x^2),1,x);
```

donen el que espereu.

Com a l'apartat anterior, treballarem amb la funció $f(x) = x^2 - 2$ i començarem les iteracions del mètode amb $x_0 = 1$.

Exercici 16.8

En una mateixa gràfica dibuixeu la gràfica de f i la de la recta tangent de f al punt $x_0 = 1$ en colors diferents. Quin és el punt d'intersecció de la recta tangent amb l'eix horitzontal? (noteu que aquest punt és el zero de la recta tangent).

El punt obtingut en l'exercici anterior l'anomenarem x_1 . I ara repetiu l'exercici anterior amb el punt x_1 en lloc del x_0 . Quin nou punt obteniu?

Noteu que després de fer tres iteracions, l'aproximació numèrica que obtenim ja és molt semblant a la que s'obté amb el Maple mitjançant la comanda `evalf(sqrt(2))`.

Exercici 16.9

El mateix que heu observat amb $f(x) = x^2 - 2$ es pot veure amb gairebé qualsevol altra funció. Feu un procediment `dibuix` que tingui com arguments una funció `f` i un punt inicial `a` i doni com a resultat el gràfic de la funció `f` i el de les tres rectes tangents obtingudes en tres iteracions consecutives del mètode de Newton partint del punt inicial `a`. (Estaria bé que, com a mínim, els colors del gràfic de `f` i de les rectes tangents fossin diferents).

La comanda `display` del paquet `plots` té l'opció `insequence=true` que permet visualitzar els diferents gràfics un després de l'altre, en comptes de tots a l'hora. Modifiqueu el procediment `dibuix` per a mostrar les diferents rectes tangents (i el gràfic de la funció) en una seqüència de dibuixos animats.

Apliqueu `dibuix` a la funció $f(x) = x^2 - 2$ amb punt inicial `a= 1` per a veure el procés que heu seguit abans per a obtenir les tres aproximacions de $\sqrt{2}$.

El problema d'aquest mètode és que en funció del punt inicial, la successió de punts obtinguda pot no convergir cap a un zero de la funció. Què passa si algun dels punts obtinguts és un extrem local (màxim o mínim)?

De fet, els punts que anem obtenint es poden descriure fàcilment de manera recursiva a partir de la fórmula de la recta tangent $r(x_{n+1}) = f(x_n) + (x_{n+1} - x_n)f'(x_n) = 0$, és a dir, $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$.

Exercici 16.10

Programem el mètode de Newton en un procediment tal com heu fet en el mètode de la bisecció. Els paràmetres del procediment són una funció f , un punt x_0 per iniciar les iteracions i el número d'iteracions. Feu que retorni una llista que contingui tots els punts obtinguts. Cal tenir en compte mentre s'executa el procediment que alguna de les rectes que apareixen pot no tenir zeros (si algun dels punts és un extrem local).

Direm que hem obtingut una precisió de n decimals si la diferència entre dues iteracions és menor de 10^{-n} .

Exercici 16.11

Modifiqueu el procediment anterior de forma que un dels paràmetres sigui la precisió desitjada enlloc del número d'iteracions. Com que aquest mètode no sempre convergeix, feu que com a molt realitzi 25 iteracions.

Exercici 16.12

Utilitzeu el mètode de Newton per a trobar un zero de la funció cosinus entre 0 i π . Preneu com a valor inicial primer 0, 0.01 i 1 i analitzeu la situació gràficament dibuixant les corresponents rectes tangents.

Exercici 16.13

Utilitzeu els procediments que heu realitzat durant la secció per a trobar solucions de l'equació $x = \cos(x)$.

Exercici 16.14

Un altre tipus de comportament patològic es pot observar en la funció $g(x) = -x^3 + 5x$.

Fixeu-vos que la funció $g(x)$ té un únic zero entre -1 i 1 (feu un gràfic). Representeu gràficament les primeres iteracions del mètode de Newton prenent com a punt inicial el 1 . Què observeu? Intenteu aproximar el zero de la funció mitjançant el mètode de la bipartició.

17 Aproximació de funcions

En aquesta secció analitzarem una qüestió d'aproximació de naturalesa diferent a la que hem treballat a la secció anterior. Donat un conjunt de dades de la forma $\{(x_1, y_1), (x_2, y_2), \dots\}$ (que podem pensar com un conjunt de punts del pla), estudiarem el problema de trobar funcions $f(x)$ de diferents tipus que passin per aquests punts, és a dir, tals que per a cada índex i es compleixi $f(x_i) = y_i$.

17.1 Polinomi interpolador

Recordem en primer lloc que Maple pot representar gràficament un núvol de punts $(x_1, y_1), \dots, (x_n, y_n)$ amb la funció `plot`.

Exercici 17.1

Definiu una llista `l` com els següents punts de \mathbb{R}^2 : $(0, 2), (1, 1), (2, 2), (3, 1), (4, 2), (5, 1), (6, 2), (7, 1), (8, 2), (9, 1)$ i $(10, 2)$. Dibuixeu en una gràfica el núvol de punts corresponent.

Donats n punts $(x_1, y_1), \dots, (x_n, y_n)$ amb $x_i \neq x_j$ si $i \neq j$ sempre existeix un polinomi p de grau menor o igual a $n - 1$ complint que $p(x_i) = y_i$ per a tots els índex i , que anomenem *polinomi interpolador*. Dins el paquet **CurveFitting** hi ha la funció `PolynomialInterpolation` que calcula polinomis interpoladors i es pot utilitzar com en l'exemple següent.

Exemple 17.1

Si volem calcular el polinomi que interpola els punts $(0, 0), (1, 1)$ i $(2, 4)$, que ha de ser de grau menor o igual a 2, ho podem fer amb Maple:

```
> with(CurveFitting);  
> ll:=[[0,0],[1,1],[2,4]]:  
   PolynomialInterpolation(ll,x);
```

Exercici 17.2

Calculeu el polinomi interpolador p als punts de l'exercici anterior. Quan val el polinomi avaluat al punt 5? I al punt 9.5? Feu la representació gràfica dels punts i el polinomi interpolador.

17.2 Spline

Una altra opció per a trobar funcions que passin per punts fixats és “*agrupar*” els punts en subconjunts més petits i calcular el polinomi de grau més petit per a cada un dels subconjunts fent que el resultat sigui una funció el més regular possible. Per exemple, podem considerar que la unió dels punts mitjançant segments es correspon a considerar els punts de dos en dos i considerar el polinomi de grau 1 (recta) que els uneix.

Dins el paquet **CurveFitting** tenim la funció **Spline** (la S en majúscules) per a definir aquestes noves aproximacions.

Exemple 17.2

Si volem unir els punts (0, 0), (1, 1) i (2, 4) amb rectes escrivim:

```
> ll := [[0,0], [1,1], [2,4]] :
      Spline(ll,x,degree=1);
```

On l'opció `degree=1` és per a definir el grau dels polinomis que estem calculant.

Exercici 17.3

Dibuixeu tres gràfiques amb els “*splines*” de graus 1, 2 i 3 de la llista de punts de l'exercici 17.1.

17.3 Exercicis

Exercici 17.4

Feu una funció (o procediment) **dibuix** que fixada una llista de punts dibuixi de colors diferents el núvol de punts, els segments que uneixen els punts, el “*spline*” de grau 2, el “*spline*” de grau 3 i el polinomi interpolador en una sola gràfica.

Exercici 17.5

Considereu la llista de punts:

```
ll i 1 := [[-9,1], [-8,1], [-7,1], [-6,1], [-5,2], [-4,3], [-3,5], [-2,8], [-1,12], [0,16],
[1,12], [2,8], [3,5], [4,3], [5,2], [6,1], [7,1], [8,1], [9,1]].
```

Dibuixeu el núvol de punts en una gràfica. Apliqueu la funció `dibuix` de l'exercici anterior a aquest núvol de punts. Que observeu?

Exercici 17.6

Definiu una funció `aproxf` que depengui de 4 paràmetres `f`, `a`, `b`, `n` que dibuixi la funció `f` entre `a` i `b` i el polinomi de grau menor o igual a `n` que interpoli a `n+1` punts equiespaiats entre `a` i `b` (inclosos).

Exercici 17.7

Apliqueu la funció `aproxf` a la funció $f(x) = \frac{1}{1+x^2}$ entre -5 i 5 i per als valors de `n` 1, 3, 6, 10 i 15. Què observeu?

17.4 Polinomi de Taylor

Un altre tipus d'aproximació polinòmica és la que s'anomena *aproximació de Taylor*. De fet l'aproximació de Taylor d'una funció és com un polinomi interpolador en el que és concentra tota la informació en un sol punt.

La comanda per a obtenir l'aproximació de Taylor de grau `n` d'una funció `f` al voltant del punt `c` és `taylor(f(x),x=c,n)` (noteu que la comanda `taylor` s'aplica a una expressió `f(x)` i no a la funció `f`).

Exemple 17.3

```
> taylor(cos(x),x=0,8);  
> taylor(ln(x),x=1,10);
```

Noteu que el resultat de `taylor` està format per una component polinòmica i una component de la forma $O((x-c)^n)$. La part polinòmica és el polinomi aproximador pròpiament dit, més endavant veurem com es pot interpretar la part $O((x-c)^n)$. Si es vol extreure el polinomi del resultat d'una comanda `taylor`, s'utilitza `convert(expr, polynom)`.

Exemple 17.4

El polinomi de Taylor de grau 5 al voltant del 0 per a la funció $f(x) = \frac{1}{1-x}$ s'obté amb

```
> expr:=taylor(1/(1-x),x=0,5);
> poli:=convert(expr, polynom);
```

Exercici 17.8

Realitzeu un procediment que, donats $f(x)$, c i n , doni com a resultat només el polinomi de Taylor de grau n de la funció $f(x)$ al voltant de c . Modifiqueu el procediment anterior per tal que retorni en una llista els polinomis de grau 1 fins a n .

Exercici 17.9

Considereu la funció $f(x) = e^x$ i el punt $c = 0$. Feu un gràfic en el que es vegin alhora el gràfic de $f(x)$ i el del seu polinomi de Taylor per a $n = 2$. Què observeu?

Repetiu l'exercici per a $f(x) = \ln(x^2 + 1)$ i per a $f(x) = \arctan(x)$.

Noteu doncs que es pot dir que el polinomi de Taylor de grau 1 ($n = 2$) és la recta tangent al gràfic de $f(x)$ en $x = c$ (això s'expressa sovint dient que *l'ordre de contacte entre els dos gràfics és 2*).

En l'exercici següent podreu veure com el concepte d'ordre de contacte s'estén a graus superiors.

Exercici 17.10

Preneu $f(x) = \ln(x^2 + 1)$ i $p(x)$ el polinomi de Taylor amb $n = 8$ de $f(x)$ al voltant de 0. Calculeu els valors de $\frac{f(x) - p(x)}{x^7}$ per a $x = \pm \frac{1}{2^n}$ (amb n des de 1 fins a 20 i amb 50 decimals com a mínim). Quin

sembla que serà el límit $\lim_{x \rightarrow 0} \frac{f(x) - p(x)}{x^7}$?

Per què es diu que s'han d'utilitzar 50 xifres com a mínim?

Com es poden obtenir els coeficients del polinomi de Taylor d'una funció a partir d'aquesta funció? Per a veure quina és aquesta relació recordeu que la comanda que dona el coeficient d'una expressió

polinòmica `expr` respecte una altra expressió `var` és `coeff(expr, var)` i que si es vol buscar el coeficient respecte `vark` també es pot fer `coeff(expr, var, k)`. Per exemple:

Exemple 17.5

```
> taylor(exp(-x^2), x=0, 9);  
> coeff(taylor(exp(-x^2), x=0, 9), x, 4);  
> taylor(sin(x), x=Pi/2, 8);  
> coeff(%, (x-Pi/2), 4);
```

Exercici 17.11

Donada $f(x) = \arctan(x)$ determineu les expressions de les derivades successives $f^{(n)}(x)$ i avalueu-les en $x = 0$ per a n des de 1 fins a 10. Feu el mateix amb $p(x)$, el polinomi de Taylor de grau 10 de $f(x)$ al voltant de 0.

Exercici 17.12

Compareu els valors que heu obtingut en l'exercici anterior amb els coeficients del polinomi de Taylor $p(x)$. (Feu el quocient entre uns i altres i si no veieu res mireu què passa si $f(x) = e^x$).

Al principi de la secció diu que el polinomi de Taylor permet aproximar una funció per un polinomi. L'exercici següent mostra com funciona aquest tipus d'aproximació.

Exercici 17.13

Feu un procediment que faci un dibuix dels gràfics (en diferents colors) d'una funció donada i dels seus polinomis de Taylor, fins a un grau n especificat en els arguments, al voltant d'un punt també donat entre els arguments. (Recordeu que, a part de poder fer un dibuix on surten tots els gràfics a l'hora, podeu fer un dibuix animat de l'evolució del polinomi de Taylor quan el grau va creixent).

Apliqueu l'anterior a la funció $f(x) = \frac{1}{(2x+1)^4}$ al voltant del punt $x = 2$ fins al grau 7.

Exercici 17.14

Les coses no van sempre tan bé com en tots els exemples anteriors. Proveu de determinar l'expressió de qualsevol polinomi de Taylor de $f(x) = e^{(-1/x^2)}$ al voltant de 0.

Tot i que `taylor` no aconsegueix fer gaire cosa amb aquest problema, hauríeu de poder determinar quin és el polinomi de Taylor de qualsevol grau per a la funció anterior al voltant de 0. I encara que es tingui el polinomi de Taylor, noteu que el resultat que s'obté no és gaire interessant!

18 Ajust a les dades. Estadística descriptiva

18.1 Ajust a les dades

Una qüestió que es planteja sovint amb dades experimentals és la següent: es realitza un experiment i es voldria saber si hi ha una relació entre dues de les seves propietats (una relació de proporcionalitat per exemple). És a dir, pot tenir interès descriure a través d'una funció la relació entre dues variables quantitatives X i Y . Posem $Y = f(X)$ on f és una funció que depèn d'uns paràmetres a_1, \dots, a_d que haurem de determinar per garantir un *bon* ajust. Noteu que en aquesta situació no és raonable esperar que l'ajust de les dades observades al model que es proposa pugui ser exacte (no es pot esperar obtenir una funció que interpoli les dades com en la secció anterior). Sempre hi haurà una certa discrepància entre les dades obtingudes i les prediccions del model, corresponent a la precisió amb la que es fan les mesures, a les simplificacions que cal fer per a poder proposar el model. . . . Per tant, a més de disposar d'observacions, $(x_1, y_1), \dots, (x_n, y_n)$ on $x_i, y_i \in \mathbb{R}$, hem de recórrer a un criteri que valori la bondat de l'ajust. El criteri que estudiarem és l'anomenat de *mínims quadrats*.

18.1.1 Llei de Hooke

Començarem plantejant un experiment concret. La llei de Hooke a la física ens diu que hi ha una relació de proporcionalitat entre el pes que es penja d'una molla i la distància que aquesta s'estira. És a dir, si x és un pes i y representa la distància que una molla s'estira en penjar-li aquell pes, existeix una constant k tal que $y = kx$. k s'anomena la constant d'elasticitat de la molla.

L'experiment que simularem consisteix en determinar aquesta constant k per a una molla concreta. Per això, s'haurien de mesurar les distàncies que la molla s'estira per a diferents pesos i s'obtidria una taula com la següent

Pes	cm
2	3.2
2.5	4
3	5
3.5	5.07
4	6.5
5	8

Si l'experiment es realitzés en condicions perfectes (mesures de precisió perfecte, sense fregament amb l'aire, . . .) les dades obtingudes serien de la forma $y_i = kx_i$ i per tant, seria fàcil determinar la constant k (tots els quocients y_i/x_i donarien k). Però com que en el moment de realitzar *realment* l'experiment aquestes condicions perfectes no es donaran, obtenim unes dades que es troben lleugerament distorsionades com es pot observar si es fa l'exercici següent.

Exercici 18.1

En uns eixos de coordenades representeu la taula de punts [Pes, cm] de l'experiment.

Per tant l'objectiu és trobar la recta de la forma $y = kx$ que millor s'ajusta a les dades obtingudes

experimentalment. Per això hem de fixar el criteri i concretar què vol dir que una recta s'ajusta millor que una altra al conjunt de dades.

El criteri a fer servir és el següent: buscarem k tal que la suma dels quadrats de les distàncies dels valors obtinguts experimentalment al valor de la recta sigui mínima. És a dir, el valor de k tal que

$$\sum_{i=1}^6 (y_i - kx_i)^2$$

pren el valor mínim. Observeu que aquesta funció mesura l'error comés en l'experiment respecte les condicions perfectes.

Exercici 18.2

Definiu una funció d tal que $d(k) = \sum_{i=1}^6 (y_i - kx_i)^2$, busqueu el valor de k pel qual d pren el valor mínim.

En una mateixa gràfica representeu els punts obtinguts a l'experiment i la gràfica de la recta $y=kx$ pel valor de k obtingut.

Aquest mètode que acabem de descriure per aproximar dades obtingudes experimentalment per funcions és el que es coneix com el mètode dels mínims quadrats.

18.1.2 El mètode dels mínims quadrats

En general, aquest mètode s'utilitza quan es vol analitzar la relació funcional entre un conjunt de dades (y_i, x_i) obtingudes experimentalment. Es suposa que hi ha un conjunt de paràmetres a_1, \dots, a_n que determina el tipus de funció f que les relaciona però no els paràmetres a_1, \dots, a_d de la funció.

El criteri de mínims quadrats proposa calcular els paràmetres a_1, \dots, a_d tals que l'expressió dels errors

$$\sum_{i=1}^n (y_i - f(x_i))^2,$$

és mínima, és a dir, la suma dels desajustos quadràtics entre els valors y_i observats i els valors $f(x_i)$ (que depenen de l'elecció dels a_i) proposats per f .

El primer pas sempre consisteix en representar gràficament en uns eixos de coordenades els punts obtinguts experimentalment com ja heu fet en la secció anterior. D'aquest tipus de diagrama se'n diu un diagrama de dispersió. Aquest tipus de gràfics es poden fer amb una comanda `plot` amb opció `style=point` però al paquet **stats** hi ha una funció dedicada específicament a aquest tipus de gràfics. Observeu el seu funcionament en el següent exemple,

Exemple 18.1

Donades les observacions següents:

$$(0.70, 0.035), (0.76, 0.025), (0.37, -0.18), (0.82, 0.045) \\ (0.29, -0.16), (0.56, -0.058), (0.42, -0.11), (0.47, -0.085),$$

el següent grup de comandes realitza la representació gràfica de les dades anteriors en un gràfic de dispersió:

```
> with(stats[statplots]):  
> X:=[0.70,0.76,0.37,0.82,0.29,0.56,0.42,0.47];  
> Y:=[0.035,0.025,-0.18,0.045,-0.16,-0.058,-0.11,-0.085];  
> scatterplot(X,Y, symbol=circle);
```

Noteu que amb `scatterplot` podem guardar en dues llistes independents els resultats de les mesures `X` i `Y` per a fer el diagrama de dispersió mentre que amb un `plot` cal fer una llista amb el valors de la forma `[[x1,y1],...]`.

Un cop fet el diagrama, el següent pas consisteix en analitzar la forma del núvol de punts obtinguts per determinar quin tipus de funció s'hi ajustaria millor.

Exercici 18.3

Quin tipus de funció creus que podríem triar per aconseguir un bon ajust a les dades de l'exemple anterior?

Exercici 18.4

Donades les observacions:

$$\begin{aligned} &(-1, 5), (-0.4, 2.5), (0.1, 2.1), (0.8, 3.98) \\ &(1.2, 6.5), (1.5, 8.8), (2.1, 15.2) \end{aligned}$$

Representeu les dades en un gràfic de dispersió. En aquest cas, quina mena de funció creus que podríem triar per ajustar a les dades?

En l'exemple 18.1 semblava adequat proposar una funció lineal $y = f(x) = a + bx$ (una recta). Ara bé, en l'exercici 18.4 sembla més adequat triar $y = f(x) = a + bx + cx^2$ (polinomi de grau 2). En ambdós casos es tracta de funcions que son lineals en els paràmetres.

Un cop determinat el tipus de funció, podem calcular els paràmetres pels quals la suma dels quadrats dels errors de mesura és mínim. En el paquet **stats**¹ hi ha la comanda **leastsquare** que calcula els paràmetres que minimitzen aquests errors. Utilitza com a paràmetres el nom de les variables, el tipus de funció i les llistes de dades experimentals.

En el exemple següent es veu com utilitzar aquesta comanda pel cas d'una recta. L'aplicarem a les dades obtingudes per l'experiment de la molla. Anomeneu **Pes** i **Cm** les dades obtingudes en l'experiment de la molla.

¹També hi ha una funció semblant en el paquet **CurveFitting** que utilitza una sintaxi lleugerament diferent

Exemple 18.2

```
> with(stats):
> fit[leastsquare[[x,y]]]([Pes,Cm]);
```

Exercici 18.5

Compareu aquest resultat amb l'obtingut a la secció 18.1.1.

Exercici 18.6

Calculeu els paràmetres de la recta que millor s'ajusta a les dades de l'exemple 18.1. En una mateixa gràfica representeu el diagrama de dispersió i la recta obtinguda.

En el cas d'ajustar altra tipus de corbes, com és el cas de l'exercici anterior on un polinomi de grau 2 de la forma $y = ax^2 + bx + c$ semblava més adequat, la sintaxi és

```
> with(stats):
> fit[leastsquare[[x,y],y=a*x^2+b*x+c,{a,b,c}]]([X,Y]);
```

Exercici 18.7

Representeu en una mateixa gràfica el diagrama de dispersió i la funció quadràtica obtinguda.

Exercici 18.8

El punt d'ebullició d'una mescla d'etanol i aigua depèn de la proporció d'etanol a la mescla. S'ha realitzat un experiment on hem obtingut les dades següents:

```
Temp=[100, 95.5, 89.0, 86.7, 85.3, 84.1, 82.7, 82.3, 81.5, 80.7, 79.8, 79.7, 79.3,
      78.74, 78.41, 78.15]
Prop=[0.0, 0.0190, 0.0721, 0.0966, 0.1238, 0.1661, 0.2337, 0.2608, 0.3273, 0.3965,
      0.5079, 0.5198, 0.5732, 0.6763,0.7472, 0.8943]
```

Si sabem que la funció que relaciona el punt d'ebullició **Temp** de la mescla i la proporció **Prop** d'etanol ha de ser de la forma $y = ae^{-14x} + bx + c$ on y és la temperatura d'ebullició i x la proporció d'etanol a la mescla. Calculeu les constants a, b, c per l'etanol.

18.2 Estadística descriptiva

El paquet **stats** no conté únicament les funcions per a determinar ajusts per mínims quadrats o per a fer gràfics de conjunts de dades, la part bàsica d'aquest paquet és un recull de subpaquets cadascun adequat per a diferents tipus d'anàlisis estadístics de dades. En aquesta secció farem un cop d'ull als subpaquets **describe**, **statsplots** i **transform**.

Exemple 18.3

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**).

```
> [1,4,6, Weight(3,2), 5$3];
```

Maple ens permet importar un fitxer de dades ja existent utilitzant **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 que es diu **"tempsvida.dat"**, i està en el directori **"c:/Mis documentos"** del nostre ordinador.²

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

ara tenim recollides en **llums** les dades que conté el fitxer.

18.2.1 Les comandes de describe

Habitualment l'anàlisi de les dades x_1, \dots, x_n obtingudes en un experiment s'inicia 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 n que apareixen en la llista de dades), **mean** (que calcula la mitjana $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ 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, $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$) i moltes altres que podreu consultar a l'ajuda de Maple a partir de **help(describe)**;

²Per a poder veure el funcionament de la comanda podeu baixar el fitxer que trobareu a <http://mat.uab.es/~gguasp/Lamp/tempsvida.dat> (on hi ha guardat el temps de vida en hores d'una mostra de llums fluorescents) i guardar-lo en qualsevol lloc del vostre sistema

Exemple 18.4

```
> 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 18.5

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

Exercici 18.9

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

18.2.2 Representació gràfica de les dades

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

18.2.3 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 18.6

```
> 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 (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).

Noteu que Maple situa el diagrama de caixa sobre la posició zero de l'eix horitzontal. Si volem que el desplaçament podem utilitzar l'opció `shift=n`, on `n` és el número d'unitats que volem que es desplaçi. 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);
```

obtindreu 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.

18.2.4 Histogrames

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

Exemple 18.7

```
> 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 dona 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 18.10

Importeu el fitxer `data3.dat` que podreu baixar de <http://mat.uab.es/~gguasp/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 o instantani. Anomeneu els tres tipus de dades com `instantani`, `escalfament` i `rapid` (recordeu com s'utilitza `importdata`).

- 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 diagrama de caixa i analitzeu el resultat.

18.2.5 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.

Per als exemples següents utilitzarem la llista **sim** que simula el llançament d'una moneda 20 cops obtinguda fent que Maple generi 20 números aleatoris enters entre 0 (=cara) i 1 (=creu) i guardant els resultats en **sim**.

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

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

Exemple 18.8

```
> 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 18.9

```
> 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 18.10

```
> transform[frequency](sim);
```

Exercici 18.11

Simuleu l'experiment que consisteix en llançar un dau 100 vegades (recordeu la comanda `rand`). Calculeu el número de vegades que ha sortit cada cara del dau i feu un gràfic on quedin reflexades aquestes dades.