SOBRE LA COMPARACIÓ ESTADÍSTICA DE CORBES EXPERIMENTALS

C.M. CUADRAS

En aquest treball es proposa un mètode d'anàlisi de la variància per a resoldre el problema de comparar dues corbes experimentals. La comparació s'estudia per els següents casos: global, parcial, per punts i per intervals.

El problema es resolt per corbes generals, però assimilades a models lineals. Es fa una anàlisi explícita dels resultats obtinguts per a comparar dues rectes de regressió.

Es fa referència a algunes aplicacions (psicología, medicina clínica, genética quantitativa). Finalment es proposen algunes generalitzacions possibles del mètode.

1. INTRODUCCIO

Es molt freqüent que el resultat d'una experiència de laboratori sigui una corba experimental, representació gràfica de la relació entre una variable observable y i una variable controlable x. Aleshores es presenta sovint el problema de cercar si hi han diferències significatives entre dues corbes experimentals, que representen el resultat de dues experiències diferents que cal comparar.

Aquest problema s'acostuma a resoldre comparant punt a punt les dues corbes mitjançant el test t de Student, fent tantes comparacions com valors té la variable controlable. Però el test t de Student és poc adequat perquè:

1) la potència del test serà en general baixa, essent possible no detectar significació a cap punt, encara que les corbes siguin verdaderament diferents.

2) encara que fitxem un nivell de significació, la probabilitat d'error de primera espècie després de diverses comparacions serà més gran.

3) quan hi ha un únic valor de y per cada valor de x, no és possible fer la comparació punt a punt per raons òbvies.

El mètode que es proposa consisteix en comparar les dues corbes globalment, mitjançant anàlisi de la variança, com una alternativa a la comparació punt a punt. S'aconsegueix amb això mantenir fitxe el nivell de significació i augmentar la potència del test. També, com veurem, es poden resoldre aspectes més matisats del problema: comparació simultània per un nombre finit de punts, comparació al llarg d'un interval, provar si hi ha paral.lelisme, etc. Inclús, en el cas de la situació 3), és possible fer la comparació per un o més punts, aplicant el mètode explicat a l'apartat 7 del present treball.

En aquest treball s'amplia, generalitza i formalitza la publicació de Cuadras i Sanchez-Turet (1977), on es proposa i comenta un mètode de comparació de corbes experimentals mitjançant la regressió polinòmica il.lustrada amb aplicacions a la psicologia (comparació del comportament de rates control amb rates desnutrides). Vegi's també Sánchez (1975).

Aplicacions a la genètica quantitativa es poden veure a Serra (1977) (comparació de línees genètiques al llarg de diverses generacions).

Una aplicació a la comparació de l'evolució de dos grups de pacients amb hepatitis, pe-

- C.M. Cuadras. Seminari Matemàtic (C.S.I.C.), Gran Via, 585. Barcelona.
- Article rebut el Febrer del 1979.

rò seguint tractaments diferents, es pot veure a Aguilar, de la Santa, Cuadras i Kern -- (1978).

Els càlculs estadístics d'aquestes aplicacions han estat realitzats per l'autor amb programes originals escrits en FORTRAN IV i processats al Laboratori de Càlcul de la Universitat de Barcelona. La finalitat d'aquest treball és fer una exposició teòrica de la metodologia estadística utilitzada.

2. MODEL PER A UNA CORBA EXPERIMENTAL

Suposem que el resultat d'una experiència sigui una corba experimental, és a dir, la representació gràfica, segons dos eixos de coordenades, de la relació entre una variable observable y i una variable controlable x, relació que és expressada per un nombre finit de parells de valors observats:

$$(x_1, y_1), \ldots, (x_n, y_n) \qquad (1)$$

Un model és una abstracció d'aquesta relació que normalment s'expressa mitjançant una - funció

$$y = F(\alpha_o, \alpha_1, \ldots, \alpha_n; x) \qquad (2)$$

en la qual intervenen també uns paràmetres - desconeguts. En models completament deterministes cal trobar valors dels paràmetres perquè (2) expressi, exactament o molt aproximadament, la relació (1). Però són més corrents en les aplicaciones els models amb part determinista F i part aleatòria e , -- essent

$$y_i = F(\alpha_o, \alpha_1, \ldots, \alpha_n; x_i) + e_i \qquad (3)$$

e_i és una variable aleatòria que representa un error o desviació aleatòria del model.

Limitaren la comparació entre dues corbes experimentals als models de tipus lineal. Per exemple:

- Recta de regressió:

$$y_i = \alpha_o + \alpha_1 x_i + e_i$$

- Polinomi de regressió:

$$y_i = \alpha_o + \alpha_1 x_i + \ldots \alpha_n x_i^n + e_i$$

- Sèrie Temporal:

$$y_i = \alpha_o + \alpha_1 x_i + \alpha_2 \sin(12x_i) + \alpha_3 \cos(12x_i) + e_i$$

També suposarem que y s'assimila a una variable aleatòria amb distribució normal. Suposarem e_i amb distribució $N(0,\sigma)$, on la desviació típica es també un paràmetre desconegut, que mesura el grau de dispersió de les dades (1) al voltant de la corba matemàtica (2). - La variable x la suposarem no aleatòria i - controlable per l'experimentador.

Donem per sabut el següent /8/ :

a) La matriu de disseny X que relaciona els valors y_i amb els valors x_i , d'acord amb el model (3). Suposem rang(X) = m + 1

b) L'estimació per mínims quadrats(estimació MQ)dels paràmetres i l'estimació de Gauss-- Markov de les funcions paramètriques estimables.

c) L'estimació de la variancia σ^2

Per a decidir si el model és adequat a les - dades, s'utilitza el mètode següent: considerem n_i observacions de y per a cada valor x de la variable controlable

$$(x_i; y_{i1}, \ldots, y_{in_i}) \qquad i = 1, \ldots, k$$

Es calcula la suma de quadrats residual

$$R_o^2 = \sum_{i=1}^{k} \sum_{h=1}^{n_i} \left[(y_{ih} - F(\hat{\alpha}_o, \hat{\alpha}_1, \ldots, \hat{\alpha}_n; x_i) \right]^2$$

essent $\hat{\alpha}_i$ les estimacions MQ dels paràmetres, i també

$$R_1^2 = \sum_{i=h}^{k} \sum_{h=1}^{n_i} (y_{ih} - \bar{y}_i)^2$$

essent

$$\bar{y} = \frac{1}{n_i} \sum_{h=1}^{n_i} y_{ih}$$

Aleshores, es demostra que l'estadístic

$$F_s = \frac{(R_1^2 - R_o^2) / (k-m-1)}{R_o^2 / (n-m-1)}$$

$(n = \sum_{i=1}^{k} n_i)$

segueix una F de Snedecor amb (k-m-1) i (n--m-1) graus de llibertat si el model (3) és vàlid. Es pot aceptar el model si la F_s calculada no és significativa /2/

3. FORMULACIO GENERAL DE LA COMPARACIO DE CORBES

Suposem ara que a més de les dades (1), tenim un altre grup de dades o una segona corba experimental

$$(\tilde{x}_1, z_1), \ldots, (\tilde{x}_{\tilde{n}}, z_{\tilde{n}}) \qquad (4)$$

observada en un altra experiència, en la qual es relaciona una variable aleatòria Z amb la variable controlable x, d'acord amb el model lineal

$$Z_i = \tilde{F}(\beta_o, \beta_1, \ldots, \beta_{\tilde{m}}; \tilde{x}_i) + e_i \qquad (5)$$

Els valors x_i i $\tilde{x}_i$ de la primera i segona experiència són els mateixos normalment, però, per a tractar el problema amb més generalitat, treballarem com si no ho fossin.

Plantejem-nos ara la següent qüestió: són diferents les dues corbes experimentals?

Si admetem que cada corba experimental està representada pel model que segueix, la resposta és: ho seràn quan coincideixin les dues corbes matematiques F i $\tilde{F}$.

Donarem solucions a aquest problema en les seccions següents.

Per a simplificar la notació. posarem $F_1=F$, $F_2=\tilde{F}$, $y_{1i}=y_i$, $y_{2i}=z_i$

Direm, a més a més, que la corba amb m+1 paràmetres es d'ordre m. Si és un polinomi,--l'ordre coincideix amb el grau.

4. COMPARACIO GLOBAL

Afirmar que les dues corbes són identiques significa aceptar la hipòtesi nul.la

$H_o^{(1)}$: $F_1 \equiv F_2$

Aquesta hipòtesi no té cap sentit si F_1 i F_2 tenen expressions analítiques diferents, o tenen diferent ordre. Una recta i una paràbola (per exemple) no poden coincidir en més de dos punts en el pla. Si les dues corbes experimentals s'ajusten bé als respectius models, de formulació matemàtica diferent, cal acceptar que les corbes experimentals son diferents.

Suposem doncs que F_1 i F_2 són del mateix tipus i que m=$\tilde{m}$. La comparació s'expressa llavors plantejant la hipòtesi lineal

$$H_o^{(2)}: \alpha_o = \beta_o\ ,\ \alpha_1 = \beta_1, \ldots, \alpha_m = \beta_m$$

Com que el nombre de punts n_i ha d'esser més gran que m+1, és fàcil de veure que $H_o^{(2)}$ és demostrable en el sentit de Roy /5/.

Siguin X_1, X_2 les matrius de disseny corresponents a la primera i a la segona corba, d'acord amb els models linelas F_1 i F_2.

Siguin també, anàlogament, y_1, y_2, els vectors d'observacions

$$y_i = (y_{i1}, y_{i2}, \ldots y_{in_i}) \qquad i=1,2$$

i $\hat{\alpha}, \hat{\beta}$ els vectors amb les estimacions MQ dels paràmetres.

La suma dels quadrats residual és, per a cada corba,

$$R_o^2(1) = (y_1 - X_1\hat{\alpha})' \ (y_1 - X_1\hat{\alpha}) = y_1' \ (y_1 - X_1\hat{\alpha})$$

$$R_o^2(2) = (y_1 - X_2\hat{\alpha})' \ (y_2 - X_2\hat{\beta}) = y_2' \ (y_2 - X_2\hat{\beta})$$

Si ara ajuntem els dos models, tindrem 2m+2 paràmetres que formen un vector $\begin{pmatrix}\alpha\\ \beta\end{pmatrix}$, un vector d'observacions $\begin{pmatrix}y_1\\ y_2\end{pmatrix}$ i una matriu de disseny

$$X_1 \otimes X_2 = \begin{pmatrix} X_1 & o \\ o & X_2 \end{pmatrix}$$

Sigui R_0^2 la suma de quadrats residual per a aquest model conjunt.

Proposició 1

$$R_0^2 = R_0^2(1) + R_0^2(2)$$

Prova:

$$R_0^2 = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}' \left(\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - X_1 \otimes X_2 \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} \right) =$$

$$= y_1' y_1 + y_2' y_2 - y_1' X_1 \hat{\alpha} - y_2' X_2 \hat{\beta} =$$

$$= y_1' (y_1 - X_1 \hat{\alpha}) + y_2' (y_2 - X_2 \hat{\beta}) = R_0^2(1) +$$

$$+ R_0^2(2)$$

A continuació suposarem que les desviacions típiques són iguals per a les dues corbes. Per això cal que l'estadístic

$$F_s = \frac{R_0^2(1) \,/\, (n_1 - m - 1)}{R_0^2(2) \,/\, (n_2 - m - 1)}$$

que té la distribució de Snedecor amb $(n_1 - m - 1)$ i $(n_2 - m - 1)$ graus de llibertat, no sigui significatiu.

Si la hipòtesi $H_0^{(2)}$ és certa, les dues corbes es confonen en una de sola, hi ha un sol vector de paràmetres $\vartheta = \alpha = \beta$ i la matriu de disseny és $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$.

La suma de quadrats residual sota la hipòtesi $H_0^{(2)}$, és

$$R_1^2 = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}' \left(\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \hat{\vartheta} \right) = y_1' (y_1 - X_1 \hat{\vartheta}) +$$

$$+ y_2' (y_2 - X_2 \hat{\vartheta})$$

essent,

$$\hat{\vartheta} = (\hat{\vartheta}_0, \ldots, \hat{\vartheta}_n)' = \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix}^{-1} (X_1'; X_2') \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

l'estimació MQ de ϑ

Proposició 2

Si H_0^2 és certa, l'estadístic

$$F_s = \frac{(R_1^2 - R_0^2(1) - R_0^2(2))/(m+1)}{(R_0^2(1) + R_0^2(2))/(n_1 + n_2 - 2m - 2)}$$

segueix la distribució F de Snedecor amb $m+1$ i n_1+n_2-2m-2 graus de llibertat

Prova: El rang de la hipòtesi lineal H_0^2 és $m+1$ i el rang de la matriu de disseny $X_1 \otimes X_2$ és $2m+2$. Com que hem demostrat que la suma de quadrats residual es $R_0^2(1)+R_0^2+(2)$ en el cas que les corbes siguin diferents, la demostració és una conseqüència del segon teorema fonamental de la teoría dels minims quadrats en model lineals /4/ (pag. 189).

Podem afirmar aleshores que les dues corbes experimentals són diferents, en sentit global, si la F_s calculada és significativa.

5. COMPARACIONS PARCIALS

De vegades interessa de comparar només alguns paràmetres de les dues corbes. Per exemple, si es tracta de dues rectes podem comparar simplement si tenen el mateix pendent, cosa que significa establir la hipòtesi $H_0: \alpha_1 = \beta_1$, però sense fer cap especificació sobre α_o i β_0.

Plantegem la hipòtesi sobre els t primers paràmetres

$$H_0^{(3)} \quad \alpha_0 = \beta_0, \alpha_1 = \beta_1, \ldots, \alpha_{t-1} = \beta_{t-1}$$

$(t < m)$

Si volguéssim comparar d'altres paràmetres, només caldria ordenar aquestos perquè fossin els primers.

Siguin

$$X_1 = (X_1^a, X_1^b) \qquad X_2 = (X_2^a, X_2^b)$$

les matrius de disseny, per a cada model, desglossades en la part X_i^a, que relaciona els t primers paràmetres, X_i^b que relaciona la resta de paràmetres. LLavors, la matriu de disseny, si $H_0^{(3)}$ és certa, serà

$$Z = \begin{pmatrix} X_1^a & X_1^b & 0 \\ X_2^b & 0 & X_2^b \end{pmatrix}$$

i l'estimació MQ del paràmetres,

$$\vartheta = (\alpha_0,\ldots,\alpha_{t-1},\alpha_t,\ldots,\alpha_n,\beta_t,\ldots,\beta_n) =$$
$$= (\beta_0,\ldots,\beta_{t-1},\alpha_t,\ldots,\alpha_n,\beta_t,\ldots,\beta_n)$$

Serà

$$\hat{\vartheta} = (Z'Z)^{-1} Z' \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \qquad (6)$$

La suma de quadrats residual sota $H_0^{(3)}$ és:

$$\bar{R}_1^2 = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}' \left(\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - Z\hat{\vartheta} \right)$$

Proposició 3

Si $H_0^{(3)}$ és certa, l'estadístic:

$$F_s = \frac{(\bar{R}_1^2 - R_0^2(1) - R_0^2(2))}{(R_0^2(1)+R_0^2(2)) \;/\; (n_1+n_2-2m-2)}$$

segueix la distribució F de Snedecor amb t i n_1+n_2-2m-2 graus de llibertat.

Prova: Es similar a l'anterior, però tenint en compte que el rang de la hipòtesi lineal $H_0^{(3)}$ és t.

Anàlogament, afirmarem que hi han diferèn-- cies significatives entre els t paràmetres si aquesta F_s és significativa.

6. COMPARACIONS PER PUNTS

Pot ésser que interessi veure només si les - dues corbes coincideixen en alguns punts, és a dir, que es tallen per alguns valors deter minats de la variable controlable.

Siguin $x_1^0,\ldots,x_t^0$, t valors comuns a les dues corbes. Plantegem la hipòtesi

$$H_0^{(4)} : F_1(\alpha_0,\alpha_1,\ldots,\alpha_n;x_i^0) =$$
$$= F_2(\beta_0,\beta_1,\ldots,\beta_n;x_i^0)$$

$i = 1,\ldots,t$

La hipòtesi es demostrable en el sentit de Roy /5/, perquè equival a afirmar que les - funcions paramètriques estimables lineals

$$\psi_i = F_1(\alpha_0,\alpha_1,\ldots,\alpha_n;x_i^0) -$$
$$- F_2(\beta_0,\beta_1,\ldots,\beta_n;x_i^0)$$

$i= 1,\ldots,t$

són iguals a 0. Aquestes funcions seran com binacions lineals del paràmetres, del tipus

$$\psi_i = (\alpha_0-\beta_0) + \rho_1(\alpha_1-\beta_1) + \ldots + \rho_n(\alpha_n-\beta_n)$$

i l'estimació de Gauss-Markov, amb variància mínima, serà :

$$\hat{\psi}_i = (\hat{\alpha}_0-\hat{\beta}_0) + \rho_1(\hat{\alpha}_1-\hat{\beta}_1) + \ldots + \rho_n(\hat{\alpha}_n-\hat{\beta}_n)$$

Aquesta variància mínima és $\text{var}(\hat{\psi}_i) = \vartheta^2_{(i)}\sigma^2$ i s'estima mitjançant

$$\hat{\sigma}^2_{\psi_i} = \vartheta^2_{(i)} \frac{(R_0^2(1)+R_0^2(2))}{(n_1+n_2-2m-2)}$$

on $\vartheta^2_{(i)}$ és una constant que depen de les - magnituds de les mostres i de les matrius de disseny X_1, X_2./2/

Per tal de decidir si les corbes es tallen en un punt de l'abcissa x_i^0 podem aplicar el S -mètode de Scheffé que permet d'establir hipòtesi simultànies sobre un conjunt de -- funcions paràmetriques estimables. /7/(pp. 68-72). Calcularem l'estadístic

$$F_s(i) = \frac{\psi_i^2}{t\hat{\sigma}^2_{\psi i}} \qquad (7)$$

que segueix la distribució F de Snedecor -- amb t i (n_1+n_2-2m-2) graus de llibertat. Si $F_s(i)$ és significativa, les corbes no es tallen en el punt de abcissa x_i^0. Si hem especificat un nivell de significació ε, es -- molt important de tenir en compte que po-- demafirmar que les corbes no es tallen en - un nombre $k \leq t$ si les $F_s(i)$ són significatives,sense que augmenti la probabilitat d'er ror de primera espècie, que es manté igual a ε. Aquesta és una propietat general del - S-mètode. També és una conseqüència del S-mètode Scheffé(/7/,pp.70-72) la següent:

Proposició 4

Cal rebutjar la hipòtesi $H_0^{(3)}$ si $F_s(i)$ és significativa per almenys un x_i^0.

Es pot donar el cas que $H_0^{(3)}$ sigui significativa però no $H_0^{(4)}$, és a dir, que les dues corbes siguin globalment diferents, però -- que en canvi, tinguin alguns punts comuns. Però rebutjar $H_0^{(4)}$ implica automàticament rebutjar $H_0^{(3)}$.

Proposició 5

Acceptar $H_0^{(4)}$ per t=n+1 punts implica acceptar $H_0^{(3)}$.

Prova: Aixo significa establir un sistema de m+1 restriccions lineals

$$\alpha_0-\beta_0+a_{i1}(\alpha_1-\beta_1)+..+a_{im}(\alpha_m-\beta_m) = 0$$

$$i = 1,...,m+1$$

que el podem interpretar com un sistema de equacions lineals homogeni amb les incognites $\alpha_0-\beta_0,\alpha_1-\beta_1,...,\alpha_n-\beta_n$. Es veu fàcilment que la solució d'aquest sistema és $\alpha_h-\beta_h=0$, $h=0,...,m$.

Fins i tot és possible també d'estudiar si dues corbes, d'ordres $m_1<m_2$, es tallen en alguns punts, determinats per alguns valors de la variable controlable. Establim doncs, la hipòtesi:

$$H_0^{(5)}: F_1(\alpha_0,\alpha_1,...,\alpha_{n1};x_i^0) =$$

$$= F_2(\beta_0,\beta_1,...,\beta_{n2};x_i^0)$$

$$i = 1,...,t$$

Utilitzarem també l'estadístic (6) per tal - de decidir si aquesta hipòtesi es pot rebutjar, però essent ara

$$\hat{\sigma}_{\psi i}^2 = \partial^2(i)\ \frac{(R_0^2(1)+R_0^2(1))}{(n_1+n_2-m_1-m_2-2)}$$

La utilització d'aquest test està limitada per la següent

Proposició 6

Si F_1 és de grau m_1 i F_2 se suposa de grau $m_2(m_1<m_2)$, acceptar $H_0^{(5)}$ implica $\beta_{m_1+1}=...=\beta_{m2}=0$. Aleshores F_2 és necessàriament de - grau m_1.

7. COMPARACIO PER INTERVALS

Sigui $[x^1,x^t]$ un interval de variació de la variable controlable. Ara volem saber si -- les dues corbes són significativament diferents per a tots els punts observats x^i dins $[x^1, x^t]$. Plantegem la hipòtesi:

$$H_0^{(6)} : F_1(\alpha_0,\alpha_1,...,\alpha_m;x^i) =$$

$$= F_2(\beta_0,\beta_1,..,\beta_m ;x^i)$$

$$x^1\leq...\leq x^i...\leq x^t$$

Veiem que no ens caldrà calcular l'estadístic (6) ni comprovar si és significatiu per a cada x^i. En efecte, sigui

$$\Delta(x^i) = \hat{\psi}_i = F_1(\hat{\alpha}_0,\hat{\alpha}_1,...,\hat{\alpha}_n;x^i) -$$

$$- F_2(\hat{\beta}_0,\hat{\beta}_1,...,\hat{\beta}_n;x^i)$$

Proposició 7

Si $\Delta(x^i)$ es creixent dins l'interval $[x^1,x^t]$ i l'estadístic

$$F_s(x^1) = \frac{\Delta^2(x^1)}{t\hat{\partial}^2_{\psi_1}}$$

és significatiu, $H_0^{(6)}$ es pot rebutjar per a tot $x^i \in [x^1, x^t]$

Prova : Es conseqüencia del fet que

$$F_s(x^1)\leq F_s(x^i)$$

$$i=2,..,t$$

Proposició 8

Si $\Delta(x^i)$ és decreixent dins l'interval - - $[x^1,x^t]$ i l'estadístic

$$F_s(x^t) = \frac{\Delta^2(x^t)}{t\,\hat{\partial}^2_{\psi_t}}$$

és significatiu, $H_0^{(6)}$ podem rebutjar per a - tot $x^i \in [x^1, x^t]$

Prova: Es conseqüència del fet que

$$F_s(x^t)\leq F_s(x^i)$$

$$i = 1,...,t-1$$

Normalment es dibuixen les corbes F_1 i F_2,- - per a fer la comparació. Aleshores és fàcil de veure si $\Delta(x^i)$ és creixent o decreixent; comprovarem llavors només la significació del primer punt o de l'últim punt de l'interval, respectivament, segons correspongui.

8. APLICACIO A LA COMPARACIO DE RECTES

Suposem que les dades s'ajusten a les rectes de regressió

$$y_{ji} = \alpha_j + \beta_j x_{ji} + e_{ji}$$

$$i = 1,\ldots,n_j \quad ; \quad j = 1,2$$

Sigui $S_{x_j y_j} = \frac{1}{n_1}\sum_i (x_{ji}-\bar{x}_j)(y_{ji}-\bar{y}_j)$ la covariància mostral i $s^2_{x_j} = 1/n \sum_i (x_{ji}-\bar{x}_j)$ la variància mostral de x. Utilitzem els mots variància i covariància per comoditat, malgrat que x no és variable aleatòria.

L'estimació MQ dels paràmetres és

$$\hat{\beta}_j = \frac{S_{x_j y_j}}{S^2_{x_j}}$$

$$\hat{\alpha}_j = \bar{y}_j - \hat{\beta}_j \bar{x}_j \qquad ; \quad j = 1,2$$

La suma de quadrats residual és

$$R_0^2 = R_0^2(1)+R_0^2(2) \text{ essent } R_0^2(j) = \sum_{i=1}^{n_j} (y_{ji}-\hat{\alpha}_j - \hat{\beta}_j x_{ji})^2 = r_j S^2_{y_j}(1-r_j^2),$$

r_j="coeficient de correlació" entre x_j, y_j. Analitzem ara cada una de les hipòtesis.

a) $H_0^{(2)}$: $\alpha_1=\alpha_2, \beta_1=\beta_2$ (comparació global)

Siguin

$$\bar{x} = \frac{1}{n_1+n_2} \sum_{j,i} x_{ji}$$

$$\bar{y} = \frac{1}{n_1+n_2} \sum_{j,i} y_{ji}$$

$$S_{xy} = \frac{1}{n_1+n_2} \sum_{j,i} (x_{ji}-\bar{x})(y_{ji}-\bar{y})$$

i anàlogament S_x^2, S_y^2, les mitjanes, covariàncies i variàncies ajuntant les dues rectes.

Indicant $\partial_1=\alpha_1=\alpha_2$, $\partial_2=\beta_1=\beta_2$ (hem fet un petit canvi de notació respecte al cas general), les estimacions MQ sota $H_0^{(2)}$ són:

$$\hat{\partial}_2 = \frac{S_{xy}}{S^2_x} \quad , \quad \hat{\partial}_1 = \bar{y} - \hat{\partial}_2 \bar{x}$$

i la suma de quadrats residual

$$R_1^2 = \sum_{j=1}^{2} \sum_{i=1}^{n_i} (y_{ji}-\hat{\partial}_1-\hat{\partial}_2 x_{ji})^2 = (n_1+n_2) S_y^2 (1-r^2)$$

essent r el "coeficient de correlació" de -- tots els punts com si les dues rectes en -- fossin una sola. Finalment , per tal de decidir sobre $H_0^{(2)}$ calcularem

$$F_s = \frac{(R_1^2 - R_0^2(1) - R_0^2(2))/\ 2}{(R_0^2(1)+R_0^2(2))\ /\ (n_1+n_2-4)}$$

amb 2 i n_1+n_2-4 graus de llibertat

b) $H_0^{(3\alpha)}$:$\alpha_1=\alpha_2$

(les rectes tenen igual el terme independent).

Indicant $\partial=\alpha_1=\alpha_2$, les equacions normals (6) són

$$\left.\begin{aligned} n\partial + \Sigma x_{1i}\beta_1 + \Sigma x_{2i}\beta_2 &= n\bar{y} \\ \Sigma x_{1i}\beta_1 + \Sigma x_{1i}^2\beta_1 &= \Sigma x_{1i}y_{1i} \\ \Sigma x_{2i}\beta_2 + \Sigma x_{21}^2\beta_2 &= \Sigma x_{2i}y_{2i} \end{aligned}\right\}$$

Calcularem seguidament

$$\bar{R}_1^2 = \sum_{j=1}^{2} \sum_{i=1}^{n_i} (y_{ji}-\hat{\partial}-\hat{\beta}_j x_{ji})^2$$

i l'estadístic

$$F_s = \frac{(\bar{R}_1^2 - R_0^2(1) - R_0^2(2))}{(R_0^2(1)+R_0^2(2))/(n_1+n_2-4)}$$

amb 1 i n_1+n_2-4 graus de llibertat

Com que $t=\sqrt{F_s}$ segueix un t de Student amb n_1+n_2-4 graus de llibertat, es podem fer proves unilaterals ($\alpha_1<\alpha_2$ ó $\alpha_1>\alpha_2$)

c) $H_0^{(3\beta)}$: $\beta_1=\beta_2$

(les rectes tenen igual pendent)

Indicant $\partial=\beta_1=\beta_2$ tenim anàlogament,

$$\left.\begin{aligned} n_1\alpha_1 \qquad\qquad\qquad + (\Sigma x_{1i})\partial &= n_1\bar{y}_1 \\ n_2\alpha_2 + (\Sigma x_{2i})\partial &= n_2\bar{y}_2 \\ (\Sigma x_{1i})\alpha_1 + (\Sigma x_{2i})\alpha_2 + (\textstyle\sum_{ji}\sum x^2_{ji})\partial &= \textstyle\sum_{ji}\sum x_{ji}y_{ji} \end{aligned}\right\}$$

Les estimacions MQ són

$$\hat{\partial} = \frac{n_2^{-1} S_{x_1y_1} + n_1^{-1} S_{x_2y_2}}{n_2^{-1} S^2_{x_1} + n_1^{-1} S^2_{x_2}}$$

$$\hat{\alpha}_1 = \bar{y}_1 - \hat{\partial}\bar{x}_1 \ , \quad \hat{\alpha}_2 = \bar{y}_2 - \hat{\partial}\bar{x}_2$$

$$\bar{R}^2 = \sum_{j=1}^{2}\sum_{i=1}^{n_i} (y_{ji} - \hat{\alpha}_j - \hat{\partial}x_{ji})^2$$

De forma similar calcularem F_s i $t = \sqrt{F_s}$

d) $H_0^{(4)} : \alpha_1 + \beta_1 x_i^0 = \alpha_2 \ \beta_2 x_i^0$

(es tallen en el punt x_i^0)

tindrem ara la funció paramètrica estimable

$$\psi_i = \alpha_1 - \alpha_2 + (\beta_1 - \beta_2)x_i^0$$

L'estimació de Gauss-Markov és

$$\hat{\psi}_i = \hat{\alpha}_1 - \hat{\alpha}_2 + (\hat{\beta}_1 - \hat{\beta}_2)x_i^0$$

i la variància mínima de $\hat{\psi}_i$ es

$$\begin{aligned} \operatorname{var}(\hat{\psi}_i) &= \operatorname{var}(\hat{\alpha}_1) + \operatorname{var}(\hat{\alpha}_2) + x_i^2 \{\operatorname{var}(\hat{\beta}_1) + \\ &+ \operatorname{var}(\hat{\beta}_2)\} + 2x_i^0 \{\operatorname{cov}(\hat{\alpha}_1, \hat{\beta}_1) + \\ &+ \operatorname{cov}(\hat{\alpha}_2, \hat{\beta}_2)\} = \end{aligned}$$

$$= \sigma^2 \left[\frac{\Sigma x_{1i}^2}{n_1 \Sigma (x_{1i} - \bar{x}_1)^2} + \frac{\Sigma x_{2i}^2}{n_2 \Sigma (x_{2i} - \bar{x}_2)^2} + (x_i^0)^2 \cdot \right.$$

$$\left\{ \frac{1}{\Sigma (x_{1i} - \bar{x})^2} + \frac{1}{\Sigma (x_{2i} - \bar{x})^2} \right\} - 2x_i^0 \left\{ \frac{\bar{x}_1}{\Sigma (x_{1i} - \bar{x}_1)^2} + \right.$$

$$\left. \left. + \frac{\bar{x}_2}{\Sigma (x_{2i} - \bar{x}_2)^2} \right\} \right] = \sigma^2 \partial^2(i)$$

L'estimació de σ^2 és

$$\hat{\sigma}^2 = (R_0^2(1) + R_0^2(2)) / (n_1 + n_2 - 4)$$

i l'estimació de $\operatorname{var}(\hat{\psi}_i)$ és $\hat{\sigma}^2_{\psi i} = \partial^2(i)\hat{\sigma}^2$

$$F_{s(i)} = \frac{\psi_i^2}{2\partial^2_{\psi_i}}$$ Es l'estadístic per tal de decidir si les dues rectes tallen en el punt d'abcissa x_i^0 amb 2 i (n_1+n_2-4) graus de llibertat

Són conseqüència de les proposicions 4 i 5:

1) Si rebutjem $H_0^{(4)}$ per a un sol punt, cal rebutjar aleshores $H_0^{(2)}$

2) Si acceptem $H_0^{(4)}$ per a dos punts, cal -- acceptar aleshores $H_0^{(2)}$

Així doncs, només té sentit aplicar el test $H_0^{(4)}$ per a 1 ó 2 punts, o fins i tot per a 3 punts en el cas que $H_0^{(4)}$ fos significatiu per a un punt i no per a l'altre; llavors podem comparar un tercer punt. Pensem que, encara que les dues rectes siguin diferents, - sino són paral.leles, sempre es tallaran en un punt. Això no contradiu pas les proposicions 4 i 5. Cal tenir en compte, però, que l'abcissa d'aquest punt pot ésser un valor - no observable de la variable controlable.

9. CONCLUSIONS

Hem vist com un planteig adequat de l'anàlisi de la variància i la teoría de models lineals ens permet comparar dues corbes experimentals.

La idea central es que la comparació (global parcial, per intervals, etc.) s'ha de fer - aprofitant tota la informació que ens donen les dades. Amb això es pretén donar una alternativa al test t de Student, que creiem poc adequat a aquest tipus de problema.

La comparació l'hem feta explícita pel cas de comparar dues rectes de regressió. Una - generalització, també molt important, consistiria en obtenir fòrmules explícites per a comparar polinomis de regressió. Deixem aquest problema per un proper treball.

Generalitzacions possibles d'aquesta contribució son les següents:

1) Comparar més de dues corbes experimentals

2) Comparar hiperplans de regressió

3) Comparar superficies de resposta.

L'extensió es pot fer sense gaires modificacions de les idees principals del treball. - Per exemple, l'estadístic per a comparar globalment k corbes de grau m (generalització de l'hipòtesi $H_0^{(2)}$) sería :

$$F_s = \frac{\{R_1^2 - \sum_{i=1}^{k} R_0^2(i)\} / (k-1)(m+1)}{\{\sum_{i=1}^{k} R_0^2(i)\} / (\sum_{i=1}^{k} n_i - k(m+1))}$$

que segueix la distribució F de Snedecor amb $(k-1)(m+1)$ i $(n_1+\ldots+n_k-k(m+1))$ graus de llibertat.

L'extensió del mètode a models no lineals ja és bastant més complicada. Cuadras i Sánchez (1977), amb el fí de comparar corbes fent - abstracció del model que segueixen, proposen descomposar la variabilitat en alçada, direcció i longitudinalment, i testar per separat la significació de cada una d'aquestes causes de variabilitat

10. BIBLIOGRAFIA

/1/ AGUILAR,J., de la SANTA,J.,CUADRAS,C.M. i KERN,A. "Zur Behandlung der chronischen hepatitis: Eine vergleichende Studie zur Wirksamkeit von Azathioprin und (+)-Cyanidanol-3". Fortschritte der Medizin, 96, 75-79, 1978.

/2/ CUADRAS, C.M. "Apuntes de Análisis de la Varianza". Bioestadística, Dpto. de Genética. Universidad de Barcelona 1978

/3/ CUADRAS, C.M. i SANCHEZ-TURET, M. "Un método de comparación entre dos curvas experimentales: una aplicación al estudio de parámetros conductuales". Revista Psicología Genética Aplicada, 32(146), 441-452. 1977

/4/ RAO, C.R. "Linear Statistical Inference and its Applications". J.Wiley, New York 2ª edición, cap. 3, secc. 3b.5, 1973.

/5/ ROY,S.N." Some aspects of Multivariate Analysis".J.Wiley, New York,Cap.12.1957

/6/ SANCHEZ-TURET, M. "Efectos de la desnutrición precoz sobre el sistema nervioso central de la rata: alteraciones en la maduración neurológica y en el funcionalismo de la actividad superior".Tesis Doctoral, U.A.B., 1975.

/7/ SCHEFFE, H. "The Analysis of Variance" J. Wiley, New York, cap. 3, Secc.3.5, 1959

/8/ SEARLE, S.R."Linear Models"J. Wiley, New York, cap.3, 1971

/9/ SERRA, L."Relación entre la variabilidad enzimática de los sistemas que controlan la Alfa-glicerofosfato-deshidrogenasa-1 y la Alcohol Deshidrogenasa y la selección por el tamaño de Drosophila melanogaster". Tesis Doctoral, U.B.1977