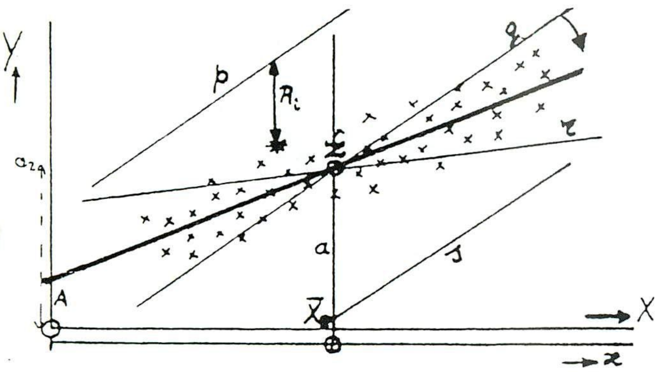
Σ Statistiek

Praktische statistiek voor klinisch farmaceutische analisten in ziekenhuis laboratoria
Lineaire of orthogonale regressie??

Hoogwaardig rekentuig en de snelle ontwikkeling van chemometrische procedures verdringen grafiekpapier en liniaal uit het laboratorium. De objectiviteit wordt hierdoor verbeterd (dit kan een nadeel zijn bij b.v. uitschieters).
Kiest men voor een bepaald rekensysteem dan is het van belang om inzicht te krijgen in de rekenprocedure.
In het navolgende worden een tweetal methoden van lineaire regressie besproken (orthogonale regressie is een vorm van lineaire regressie). Bij conventionele lineaire regressie volgens de 1e orde wordt de lijn beschreven door de vergelijking Y = A + Bx.
A is de afsnijding van de Y-as en B is de richtingscoefficiënt van de lijn. De parameters A en B hebben een (onbekende) ware waarde en de opdracht is A en B zo goed mogelijk te benaderen uit een waarnemingsreeks. Daartoe doen we n waarnemingen. We zetten deze grafisch uit en krijgen zo n punten (Xi; Yi).



In de meeste praktijkgevallen is een van de twee variabelen X en Y veel nauwkeuriger bekend dan de andere. De meest nauwkeurige variabele kiezen we altijd als de X-variabele.
Bij lineaire regressie nemen we dan ook aan dat;

-de fluctuaties uitsluitend in de Y zitten en dat de fouten in de X verwaarloosbaar zijn;

-de standaardafwijking constantis d.w.z. onafhankelijk van de waarde van Y;

-de metingen elkaar niet beïnvloeden.
Voor de berekening van de lijn wordt door het gros der rekentuigen gebruik gemaakt van de methode der kleinste kwadraten.
De kwadratensom Q wordt berekend als de som van de kwadraten van alle afstandjes (Ri) tot de lijn. Q is een functie van A en B.

$$Q(A\ B) = Ri^2 + R2^2 + R3^2 + \ldots\ldots\ldots\ldots Rn^2$$

Q gaat door een minimum als we de lijn P evenwijdig aan zichzelf opschuiven naar n. Het minimum in Q als functie van A wordt bereikt als de lijn door het zwaartepunt Z van de wolkpunten gaat. Deze minimumwaarde van Q kunnen we echter nog verder verlagen door de lijn te roteren in de richting van Z. Het rotatiepunt bij conventionele lineaire regressie is A en bij deze draaiing verlaten we het zwaartepunt Z en daarbij tevens de beste positie qua hoogte. Op deze wijze kunnen we B niet onafhankelijk van A aanpassen. Bij orthogonale regressie wordt de lijn altijd door het zwaartepunt getrokken. De vergelijking van de rechte lijn gaat dan over in:

$$Y = az + B\ (xi - xz)$$

az = de hoogte van het zwaartepunt Z op de Y-as.
Het voordeel van deze methode is dat az en B onafhankelijk van elkaar zijn.
Principieel is de orthogonale regressie methode dus altijd beter. Praktisch zal het verschil tussen de conventionele lineaire en orthogonale regressie indien de spreiding in de X inderdaad zeer klein is t.o.v. de Y, nihil zijn. Indien de spreiding in X niet te verwaarlozen is t.o.v. de spreiding in Y, zoals bij de vergelijking tussen 2 methoden dan is er wel verschil en heeft de orthogonale regressie de voorkeur.
Jack van der Heijden.

ALS LIJNEN KONDEN SPREKEN EN MEETPUNTEN KONDEN DENKEN DAN ZOU ER VEEL MEER GEVOEL IN ZITTEN.