

PEMODELAN DALAM ANALISIS DATA UJI KLINIK (I): MODEL REGRESI LINEAR DAN REGRESI LOGISTIK

Johan Harlan

Pusat Studi Informatika Kedokteran
Universitas Gunadarma
Jl. Margonda Raya No. 100 Depok

Abstrak

Dari sudut pandang Statistika, pemodelan dalam pengolahan data penelitian eksperimental seperti pada halnya uji klinik tidaklah dianjurkan. Randomisasi apabila dilaksanakan sebagaimana mestinya dianggap telah mengoptimalkan keseimbangan distribusi variabel prediktor manapun antar kelompok perbandingan di luar faktor perlakuan utama (pengobatan dalam uji klinik), sehingga peningkatan validitas dan presisi yang menjadi tujuan utama pada pengolahan data penelitian observasional menjadi kurang relevan di sini. Dalam kenyataannya penggunaan pemodelan dalam pengolahan data uji klinik masih relatif sering ditemukan, walaupun interpretasinya perlu ditanggapi secara berbeda dengan interpretasi terhadap hasil analisis multivariabel pada penelitian observasional secara umumnya.

Kata Kunci: Uji klinik, pemodelan, model regresi linear, model regresi logistik

PENDAHULUAN

Pemodelan merupakan salah bentuk analisis multivariabel dalam pengolahan dan analisis data hasil penelitian. Dalam uji klinik, analisis multivariabel digunakan untuk menilai hubungan antara jenis pengobatan dan respons dengan mengendalikan kovariabel (faktor prognostik yang dianggap mempengaruhi respons pengobatan).

Secara teoritis, terdapat tiga tujuan yang mungkin dicapai menggunakan pemodelan (dan bentuk analisis multivariabel lainnya) pada analisis data hasil penelitian, yaitu peningkatan validitas (melalui pengendalian konfounding), peningkatan presisi, serta pendeteksian interaksi antara jenis pengobatan dengan kovariabel. Walaupun demikian, dalam suatu uji klinik dengan randomisasi yang terlaksana sebagaimana mestinya, peningkatan validitas serta presisi demikian umumnya dapat dikatakan tidak memiliki nilai klinis, bahkan untuk pengendalian kovariabel yang ditandai dengan adanya ketidakseimbangan yang sangat jelas antar kelompok pengobatan sekalipun. Demikian pula halnya dengan pendeteksian interaksi, terdapatnya jumlah kemungkinan interaksi yang sangat besar dalam suatu uji klinik akan menimbulkan kesulitan dalam pengujian maupun penginterpretasiannya. Berperannya efek

perbandingan ganda di sini menyebabkan adakalanya interaksi yang sangat bermakna secara statistik pun setelah dinilai secara klinis hanya dapat dijelaskan sebagai peris-tiwa kebetulan (kesalahan tipe I) belaka.

Dengan demikian, dalam suatu uji klinik dengan randomisasi yang terlaksana sebagaimana mestinya, pemodelan (serta bentuk analisis multivariabel lainnya) umumnya hanya dianggap sebagai suatu bentuk analisis sekunder dengan tujuan utama untuk lebih memahami patogenesis serta mekanisme perjalanan penyakit dalam konteks hubungan antara pengobatan yang dipelajari dengan respons, serta meletakkan landasan ilmiah bagi penelitian lebih lanjut (Meier, 1983).

MODEL REGRESI LINIER

Model regresi linear yang akan dibahas di sini adalah model untuk respons kategorik dengan skala dikotomi, sedangkan untuk variabel bebas dapat digunakan data numerik maupun data kategorik (dalam bentuk variabel buatan). Model regresi linear dapat digunakan untuk analisis data proporsi yang tidak sesuai hasil rancangan antar-subjek dengan randomisasi lengkap apabila efek variabel bebas (jenis pengobatan beserta kovariabelnya) diasumsikan bersifat aditif.

Model yang digunakan di sini adalah:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i \quad (1)$$

dengan:

Y_i : respons untuk subjek ke- i dalam skala dikotomi, misalnya $Y_i = 0$ menyatakan subjek ke- i yang tidak sembuh; $Y_i = 1$ menyatakan subjek ke- i yang sembuh

X_{ij} : jenis pengobatan pada subjek ke- i , misalnya $X_{i1} = 0$ menyatakan subjek ke- i dengan pengobatan kontrol (obat B); $X_{i1} = 1$ menyatakan subjek ke- i dengan pengobatan uji (obat A)

X_{ij} : kovariabel yang akan dikendalikan, dapat dinyatakan dalam bentuk data numerik ataupun kategorik (variabel buatan); $j = 2, 3, \dots, p$

β_j : parameter populasi (koefisien regresi); $j = 0, 1, \dots, p$

ε_i : galat acak pada subjek ke- i

Untuk menyatakan interaksi antara efek jenis pengobatan X_1 dan kovariabel X_j terhadap respons, dapat ditambahkan suku $\beta_k X_{i1} X_{ij}$ pada model 1 di atas.

Model 1 dapat pula dituliskan dalam bentuk penyajian matriks:

$$Y = X\beta + \varepsilon \quad (2)$$

dengan:

Y : vektor respons

β : vektor parameter

X : matriks variabel bebas

ε : vektor galat acak

Misalkan vektor $b = \beta$ menyatakan perkiraan untuk parameter pada model 1 yang akan dihitung:

$$b = [b_0 b_1 \dots b_p]'$$

maka perkiraan kuadrat terkecil untuk β_j adalah nilai-nilai b_j yang meminimum-kan jumlah

kuadrat:

$$Q = \sum_{i=1}^n [Y_i - \hat{Y}_i]^2 \quad (3)$$

dengan:

$$Y = f(\mathbf{b}) = b_0 + b_1 X_{i1} + b_2 X_{i2} + \dots + b_p X_{ip}$$

Dalam bentuk penyajian matriks di-peroleh estimator kuadrat terkecil untuk parameter populasi (yang sekaligus merupakan estimator kemungkinan maksimum):

$$b = (X'X)^{-1} X'Y \quad (3.a)$$

Metode ini hanya tepat untuk di-gunakan pada model regresi linear yang memenuhi beberapa asumsi tertentu, antara lain suku galat acak ε_i berdistribusi normal dengan rata-rata 0 dan ragam σ^2 yang konstan (asumsi homoskedastisitas).

Pada model 1 di atas dengan respons kategorik dikotomi asumsi ini tidak terpenuhi, sehingga untuk perkiraan parameternya digunakan metode kuadrat terkecil tertimbang, dengan jumlah kuadrat yang harus diminimum-kan (Neter & Wasserman, 1974) adalah:

$$Q = \sum_{i=1}^n W_i [Y_i - \hat{Y}_i]^2 \quad (4)$$

Sebagai penimbang digunakan:

$$W_i = \frac{1}{\hat{Y}_i(1 - \hat{Y}_i)} \quad (4.a)$$

Dalam pelaksanaannya, nilai-nilai \hat{Y}_i tidak diketahui, sehingga W_i tak dapat di-tentukan secara langsung. Karena itu perkiraan parameter dilakukan dalam dua tahap. Mula-mula dilakukan perkiraan parameter sementara dengan metode kuadrat terkecil biasa. Berdasarkan nilai penduga sementara ini dapat dihitung nilai-nilai \hat{Y}_i . Selanjutnya proses perkiraan parameter diulangi kembali dengan menggunakan metode kuadrat terkecil tertimbang.

Misalkan W menyatakan matriks berikut:

$$W = \begin{bmatrix} W_1 & 0 & 0 & \dots & \dots & 0 \\ 0 & W_2 & 0 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \dots & W_n \end{bmatrix}$$

maka dalam bentuk penyajian matriks diperoleh perkiraan parameter:

$$b = (X' W X)^{-1} X' W Y \quad (4.b)$$

Perhatikan bahwa:

$$E(Y) = \sum Y_i P(Y_i) = (0)P(Y_i = 0) + (1)P(Y_i = 1) \\ = P(Y_i = 1) = p \quad (5)$$

menyatakan probabilitas terjadinya respons, yaitu:

$$E(Y) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}$$

karena itu, dalam aplikasi model 1 di atas untuk prediksi, hanya boleh digunakan nilai X_{ij} ; $j = 0, 1, \dots, p$ sedemikian rupa sehingga:

$$0 \leq E(Y) = p \leq 1$$

Untuk pembahasan lebih lengkap mengenai model regresi linear dengan respons dikotomi, lihat Neter & Wasserman (1974), dengan contoh yang diberikan berupa model dengan variabel bebas numerik. Walaupun demikian, model regresi linear ini dapat pula digunakan pada keadaan dengan variabel terikat dan bebas keseluruhannya merupakan data kategorik (Grizzle, Starmer, Koch; 1969), yang juga menggunakan metode kuadrat terkecil tertimbang.

Contoh 1:

Misalkan dalam suatu uji klinik, R (respons) menyatakan variabel terikat, T (perlakuan) menyatakan jenis pengobatan, dan C menyatakan kovariabel (faktor prognostik) dengan kesemuanya merupakan data kategorik dalam skala dikotomi. Maka model regresi linear tanpa interaksi adalah:

$$R_i = \alpha + \beta T_i + \gamma C_i + \varepsilon_i \quad (6)$$

sedangkan model dengan interaksi dapat dinyatakan dalam bentuk:

$$R_i = \alpha + \beta T_i + \gamma C_i + \delta T_i C_i + \varepsilon_i \quad (7)$$

Perhatikan model 6. Untuk kelompok pengobatan uji (obat A) nilai $T_i = 1$, sedangkan untuk kelompok pengobatan kontrol (obat B) nilai $T_i = 0$, sehingga

$$\text{diperoleh: } P(R_i = 1 | T_i = 1; C_i) = \alpha + \beta(1) + \gamma C_i$$

$$= \alpha + \beta + \gamma C_i$$

$$P(R_i = 1 | T_i = 0; C_i) = \alpha + \beta(0) + \gamma C_i$$

$$= \alpha + \gamma C_i$$

Perbedaan risiko kejadian disesuaikan (perbedaan kejadian kumulatif disesuaikan) adalah:

$$aCID = P(R_i = 1 | T_i = 1; C_i) - P(R_i = 1 | T_i = 0; C_i) \\ = (\alpha + \beta + \gamma C_i) - (\alpha + \gamma C_i) = \beta \quad (8)$$

dan perkiraannya berdasarkan data sampel adalah:

$$a\hat{C}ID = \hat{\beta} = b \quad (8.a)$$

Untuk menilai interaksi perhatikan model 7. Kovariabel C terdiri dari dua taraf, yaitu $C_i = 1$ untuk stratum I dan $C_i = 0$ untuk stratum II, sehingga diperoleh:

$$P(R_i = 1 | T_i = 1; C_i) = \alpha + \beta(1) + \gamma(1) + \delta(1)(1)$$

$$= \alpha + \beta + \gamma + \delta$$

$$P(R_i = 1 | T_i = 0; C_i) = \alpha + \beta(0) + \gamma(1) + \delta(0)(1)$$

$$= \alpha + \gamma$$

Perbedaan kejadian kumulatif pada stratum I adalah:

$$CID = P(R_i = 1 | T_i = 1; C_i = 1) - P(R_i = 1 | T_i = 0; C_i = 1)$$

$$= (\alpha + \beta + \gamma + \delta) - (\alpha + \gamma)$$

$$= \beta + \delta \quad (9.a)$$

dan perkiraannya adalah:

$$\hat{C}ID = \hat{\beta} + \hat{\delta} = b + d \quad (9.b)$$

Dengan cara yang sama diperoleh :

$$CID_2 = I(R=1|T_1=1;C_1=0) - I(R=1|T_1=0;C_1=0) = (\alpha + \beta) - \alpha = \beta \quad (9.c)$$

dan perkiraannya:

$$\hat{CID}_2 = \hat{\beta} = b \quad (9.d)$$

Besar interaksi (dalam skala aditif) di sini adalah :

$$I_a = CID_1 - CID_2 = (\beta + \delta) - \beta = \delta \quad (10)$$

Perkiraan adalah :

$$\hat{I}_a = \hat{\delta} = d \quad (10.a)$$

Perhatikan bahwa nilai-nilai α , β , dan δ pada model 7 ini tidak sama dengan pada model 6.

Uji hipotesis yang dianjurkan di sini adalah uji khi-kuadrat residual untuk menilai kesesuaian model (untuk model tereduksi), serta uji Wald yang menghasilkan nilai p bagi masing-masing koefisien regresi (untuk uji interaksi dan asosiasi). Statistik pengujian untuk uji khi-kuadrat residual adalah:

$$\chi^2 = \sum \frac{(\text{Nilai-nilai pengamatan} - \text{nilai-nilai prediksi})^2}{\text{Nilai-nilai prediksi}} \quad (11)$$

yang berdistribusi khi kuadrat dengan derajat bebas satu (untuk model tereduksi 6). Untuk model 7 tidak dilakukan uji statistik karena nilai prediksi pada model jenuh akan tepat sama dengan nilai pengamatan. Statistik pengujian untuk uji Wald adalah:

$$\chi = Z = \frac{\hat{\beta}}{SE \hat{\beta}} \quad (12)$$

yang berdistribusi normal standar. Hasil uji hipotesis demikian umumnya akan langsung diperoleh pada keluaran program statistik komputer.

MODEL REGRESI LOGISTIK

Aplikasi model regresi linear dalam analisis data hasil uji klinik mempunyai kelemahan, antara lain dengan dimungkinkannya memperoleh perkiraan proporsi kejadian respons yang bernilai

negatif ataupun lebih besar daripada satu, walaupun hal ini mungkin lebih disebabkan oleh metode yang digunakan untuk mengperkirakan koefisien regresi daripada ketidaksesuaian model linear itu sendiri (Kleinbaum, 1982).

Karena itu, untuk analisis data proporsi yang tidak sesuai hasil rancangan an-tar-subjek dengan randomisasi lengkap, model regresi logistiklah yang umumnya lebih disukai. Selain itu, model ini dapat pula digunakan untuk analisis data hasil rancangan an-tar-subjek dengan kesesuaian. Dalam model ini, hubungan antar variabel bebas diasumsikan bersifat multiplikatif.

Model yang digunakan di sini adalah:

$$Y_i = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}))}$$

atau : $\text{logit } Y_i = \ln \frac{Y_i}{1 - Y_i}$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} \quad (13)$$

dengan :

$Y_i = p_i$: probabilitas terjadinya respons pada subjek ke- i . Pada analisis data sampel, Y_i adalah respons untuk subjek ke- i , yang dinyatakan dalam skala dikotomi.

X_{i1} : jenis pengobatan pada subjek ke- i , dinyatakan dalam skala dikotomi.

X_{ij} : kovariabel yang akan dikendalikan, dapat dinyatakan dalam bentuk data numerik ataupun kategorik (variabel buatan); $j = 2, 3, \dots, p$

β_j : parameter populasi (koefisien regresi); $j = 0, 1, \dots, p$

Perkiraan parameter yang dianggap terbaik untuk model regresi logistik dilakukan dengan metode kemungkinan maksimum, yang umumnya dikerjakan dengan komputer karena memerlukan proses iteratif dalam perhitungannya (tidak dibahas di sini). Untuk data yang tidak sesuai, umumnya digunakan metode kemungkinan maksimum tak bersyarat, sedangkan untuk data

yang sesuai dianjurkan penggunaan metode kemungkinan maksimum bersyarat. Untuk pembahasan lebih lengkap mengenai model regresi logistik, lihat Kleinbaum (1982).

Contoh 2:

Lihat kembali contoh 1. Misalkan hubungan antara jenis pengobatan T dan faktor prognostik C bersifat multiplikatif, maka model yang digunakan, masing-masing tanpa dan dengan interaksi adalah

$$\text{logit } R_i = \alpha + \beta T_i + \gamma C_i \quad (14)$$

dan

$$\text{logit } R_i = \alpha + \beta T_i + \gamma C_i + \delta T_i C_i \quad (15)$$

Perhatikan bahwa nilai-nilai parameter $\alpha, \beta, \gamma,$ dan δ pada tiap model tidak sama.

Lihat model 14. Untuk kelompok pengobatan uji (obat A) berlaku:

atau:

$$\frac{P(R_i = 1 | T_i = 1; C_i)}{P(R_i = 0 | T_i = 1; C_i)} = \exp(\alpha + \beta + \gamma C_i)$$

sedangkan untuk kelompok pengobatan kontrol (obat B) berlaku :

$$\begin{aligned} \text{logit } R_i &= \ln \frac{P(R_i = 1 | T_i = 0; C_i)}{P(R_i = 0 | T_i = 0; C_i)} \\ &= \alpha + \beta(0) + \gamma C_i = \alpha + \gamma C_i \end{aligned}$$

$$\text{atau: } \frac{P(R_i = 1 | T_i = 0; C_i)}{P(R_i = 0 | T_i = 0; C_i)} = \exp(\alpha + \gamma C_i)$$

Rasio imbangannya adalah:

$$\begin{aligned} aOR &= \frac{P(R_i = 1 | T_i = 1; C_i)}{P(R_i = 0 | T_i = 1; C_i)} \bigg/ \frac{P(R_i = 1 | T_i = 0; C_i)}{P(R_i = 0 | T_i = 0; C_i)} \\ &= \frac{\exp(\alpha + \beta + \gamma C_i)}{\exp(\alpha + \gamma C_i)} = \exp \beta \end{aligned} \quad (16)$$

dan perkiraannya adalah:

$$a\hat{O}R = \exp \hat{\beta} = \exp b \quad (16.a)$$

Untuk menilai interaksi, perhatikan model 15. Kovariabel C terdiri dari dua taraf,

$$\begin{aligned} \text{logit } R_i &= \ln \frac{R_i}{1 - R_i} \\ &= \ln \frac{P(R_i = 1 | T_i = 1; C_i)}{1 - P(R_i = 1 | T_i = 1; C_i)} \\ &= \ln \frac{P(R_i = 1 | T_i = 1; C_i)}{P(R_i = 0 | T_i = 1; C_i)} \\ &= \alpha + \beta(1) + \gamma C_i \\ &= \alpha + \beta + \gamma C_i \end{aligned}$$

yaitu $C_i = 1$ untuk stratum I dan $C_i = 0$ untuk stratum II, sehingga untuk stratum I:

$$\ln \frac{P(R_i = 1 | T_i = 1; C_i = 1)}{P(R_i = 0 | T_i = 1; C_i = 1)} = \alpha + \beta(1) + \gamma(1) + \delta(1)(1) = \alpha + \beta + \gamma + \delta$$

dan:
$$\frac{P(R_i = 1 | T_i = 1; C_i = 1)}{P(R_i = 0 | T_i = 1; C_i = 1)} = \exp(\alpha + \beta + \gamma + \delta)$$

$$\ln \frac{P(R_i = 1 | T_i = 0; C_i = 1)}{P(R_i = 0 | T_i = 0; C_i = 1)} = \alpha + \beta(0) + \gamma(1) + \delta(0)(1) = \alpha + \gamma$$

dan
$$\frac{P(R_i = 1 | T_i = 0; C_i = 1)}{P(R_i = 0 | T_i = 0; C_i = 1)} = \exp(\alpha + \gamma)$$

$$OR_1 = \frac{P(R_i = 1 | T_i = 1; C_i = 1)}{P(R_i = 0 | T_i = 1; C_i = 1)} \bigg/ \frac{P(R_i = 1 | T_i = 0; C_i = 1)}{P(R_i = 0 | T_i = 0; C_i = 1)}$$

$$= \frac{\exp(\alpha + \beta + \gamma + \delta)}{\exp(\alpha + \gamma)} = \exp(\beta + \delta) \tag{17.a}$$

dan perkiraannya :

$$\hat{OR}_2 = \exp(\hat{\beta}) = \exp(b) \tag{17.d}$$

Perkiraan adalah:

$$\hat{OR}_1 = \exp(\hat{\beta} + \hat{\delta}) = \exp(b + d) \tag{17.b}$$

Dengan cara yang sama untuk stratum II diperoleh:

$$OR_2 = \exp(\beta) = \exp \delta \tag{17.c} \tag{18.a}$$

atau:

$$\ln I_m = \ln OR_1 - \ln OR_2 = (\beta + \delta) - \beta = \delta \tag{18.b}$$

dan perkiraannya adalah:

$$\hat{I}_m = \exp \hat{\delta} = \exp d \tag{18.c}$$

atau

$$\ln \hat{I}_m = \hat{\delta} = d \tag{18.d}$$

Untuk uji hipotesis $H_0: OR = 1$ atau H_0 , tidak ada interaksi antara jenis pengobatan T dan kovariabel C , yang dianjurkan adalah uji rasio *likelihood*, yang umumnya juga dikerjakan dengan komputer. Misalkan dimiliki model 1 (tanpa suku

Perhatikan bahwa interaksi yang dimaksud di sini seandainya ada ($OR_1 \neq OR_2$) ialah interaksi dalam skala multiplikatif. Besarnya interaksi adalah:

$$I_m = OR_1 / OR_2$$

βT , untuk uji perbandingan efek pengobatan atau tanpa suku $\delta T, C_i$ untuk uji interaksi) dengan k parameter dan fungsi *likelihood* $L1$, serta model 2 (dengan suku $\beta T, \delta T, C_i$ sesuai hipotesis yang diuji) dengan $(k + 1)$ parameter dan fungsi *likelihood* $L2$, maka statistik pengujinya adalah:

$$G = -2 \ln (L1/L2) \tag{19}$$

yang berdistribusi khi-kuadrat dengan derajat bebas satu. Perkiraan fungsi *likelihood* $L1$ dan $L2$ diperoleh dari keluaran program komputer. Uji Wald juga dapat digunakan di sini, walaupun kurang dianjurkan penggunaannya.

Perhatikan bahwa berbeda dengan analisis sederhana ataupun analisis strati-fikasi, pada analisis multivariabel dengan pe-modelan uji

hipotesis $H_0: CID = 0$ tidak sama dengan uji hipotesis $H_0: OR = 1$, karena keduanya menggunakan model yang berbeda (model regresi linear untuk CID dan model regresi logistik untuk OR) dengan asumsi-asumsi yang berbeda pula. Juga pada penggunaan program komputer, apabila variabel buatan yang digunakan bernilai -1 dan $+1$ (bukan 0 dan 1 seperti pada pembahasan di atas), rumus yang diberikan di sini harus disesuaikan lagi.

PENUTUP

Pemodelan dalam analisis data hasil uji klinik adalah penggunaan model matematika untuk menggambarkan hubungan antara respons dengan himpunan variabel bebas, yang terdiri dari jenis pengobatan serta satu atau lebih kovariabel yang akan dikendalikan.

Untuk respons yang bersifat dikotomi serta jenis pengobatan yang bersifat ka-tegorik dapat digunakan model regresi linear untuk data kategorik atau model regresi logistik. Model regresi linear digunakan apa-bila hubungan antar variabel bebas diasum-sikan bersifat aditif, sedangkan model regresi logistik digunakan pada hubungan antar variabel bebas yang bersifat multiplikatif. Per-kiraan parameter dilakukan dengan metode kuadrat terkecil tertimbang pada model regresi linear untuk data kategorik dan metode kemungkinan maksimum pada model regresi logistik.

DAFTAR PUSTAKA

- Anderson, S, et al. **Statistical Methods for Comparative Studies: Techniques for Bias Reduction**. New York: John Wiley & Sons, 1980.
- Dobson, AJ. **Introduction to Statistical Modelling**. London: Chapman and Hall, 1983.
- Kleinbaum, DG, et al. **Epidemiologic Research: Principles and Quantitative Methods**. New York: Van Nostrand Reinhold Company, 1982.
- Meier, P. **Statistical Analysis of Clinical Trials**. In: Shapiro, SH & Louis, TA (eds), **Clinical Trials: Issues and Approaches**. New York: Marcel Dekker, Inc, 1983.
- Meinert, CL. **Clinical Trials: Design, Conduct, and Analysis**. New York: Oxford University Press, 1986.
- Neter, J & Wasserman, W. **Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs**. Homewood, Illinois: Richard D Irwin, Inc, 1977.
- Parker, JC. **SAS / STAT Guide for Personal Computers, Version 6 Edition**. Cary, NC: SAS Institute Inc, 1987.