

ภาคผนวก ข

**สรุปคำสั่งของกระบวนการ REGRESSION FACTOR และ CLUSTER
รายละเอียดคำสั่งของกระบวนการ REGRESSION**

สรุปคำสั่งของกระบวนการความ REGRESSION

COMMAND นี้มี subcommand ที่จำเป็น 3 คำสั่งคือ VARIABLES DEPENDENT และ METHOD นอกนั้นเรียกใช้ตามความจำเป็น ซึ่งแต่ละ SUBCOMMAND มี KEYWORD ให้เลือกใช้ตามความต้องการซึ่งอาจผิดแผกกันตามวัตถุประสงค์ของงานต่าง ๆ ดังนี้

/ VARIABLES ให้เขียนต่อจากชื่อกระบวนการความ REGRESSION เป็น

REGRESSION VARIABLES = ชื่อตัวแปรต่าง ๆ ทั้งหมด

/ DEPENDENT = ชื่อตัวแปรตาม

/ METHOD = FORWARD BACKWARD STEPWISE ENTER (varlist) REMOVE (varlist) TEST (varlist)

subcommand นี้ต้องสั่งต่อจาก /DEPENDENT ในทันที เราสามารถสั่งได้ในลักษณะ multiple subcommand ได้ (ดูตัวอย่าง 2)

/ CRITERIA = DEFAULTS PIN (value) POUT (value) FIN (value) FOUT (value) TOLERANCE (value) MAXSTEP (n)

โดยปกติถึงไม่สั่ง subcommand นี้โปรแกรมก็เรียก keyword DEFAULT ทำงานซึ่งจะกำหนด PIN (.05) POUT (.10) และ TOLERANCE (.01) เราต้องวาง /CRITERIA หลัง VARIABLES และก่อน /DEPENDENT (ดูตัวอย่าง 3)

/ STATISTICS = DEFAULTS ALL R ANOVA CHA BCOV XTX COND COEFF OUTS ZPP CI SES TOLER LINE HISTORY END

คำสั่ง /STATISTICS ต้องวางอยู่ก่อน /DEPENDENT ถ้าไม่สั่งก็จะวิเคราะห์เฉพาะ default คือ R ANOVA COEFF OUTS (ดูตัวอย่าง 4)

/ ORIGIN ใช้สำหรับสั่งให้สมการมี intercept term (/NOORIGIN) หรือไม่มี intercept term (/ORIGIN) โดยปกติ /NOORIGIN จะเป็น default (คือไม่ต้องระบุ) ถ้าจะมีคำสั่ง /ORIGIN ให้สั่งไว้ระหว่าง /VARIABLES กับ /DEPENDENT (ดูตัวอย่าง 5)

/ SELECT = ชื่อตัวแปร ความสัมพันธ์
ค่าความสัมพันธ์คือ EQ NE LE LT GT GE ใช้สำหรับสร้างสมการเฉพาะ case
ที่เราต้องการ เช่น เฉพาะเพศ เฉพาะอายุใดอายุหนึ่ง การสั่ง **/SELECT**
ให้สั่งต่อเนื่อง ก่อนหรือหลัง ทันทันใดกับ **/VARIABLES** (ดูตัวอย่าง 6)

/ MISSING = LISTWISE PAIRWISE MEANSUBSTITUTION INCLUDE
ใช้สั่งให้จัดการกับข้อมูลที่สูญหาย ถ้าไม่สั่ง **/MISSING** โปรแกรมจะตัด
ค่าสังเกตที่มี missing value ทิ้ง (หมายความว่า LISTWISE
เป็น default) ถ้าสั่ง **/MISSING** ให้สั่งต่อเนื่องทันที(ก่อนหรือหลัง)
/VARIABLES

DESCRIPTIVES = DEFAULTS MEAN STDDEV VARIANCE CORR SIG BAD CORR
COV XPROD N ALL
ถ้ากำหนด **/DESCRIPTIVES** แต่ไม่ระบุ keyword จะทำ default
คือ MEAN STDDHV CORR (ดูตัวอย่าง 7)

Subcommand ต่อไปนี้ใช้สำหรับวิเคราะห์ residual คือ **/RESIDUALS /CASEWISE**
/PARTIALPLOT /SCATTERPLOT Subcommand เหล่านี้ต้องต่อท้ายคำสั่ง **/METHOD**
ล่าสุดโดยทันที

/ RESIDUALS = DEFAULTS HISTOGRAM(tempvars) NORMPROB(tempvars)
SIZE(plotsize) OUTLIERS(tempvars) DURBIN ID(varname)
POOLED
ถ้าสั่ง **/RESIDUALS** เฉย ๆ โดยไม่ระบุ keyword จะทำ default
ใน default มี keyword ต่าง ๆ ซึ่งทำการวิเคราะห์ให้เองถ้าไม่ระบุ
key word ใด ๆ คือ HISTOGRAM(ZRESID) NORMPROB(ZRSID)

SIZE(SMALL) PLOTS(ZRESID) และ DURBIN SPSS จะถือว่าค่า
ค่าพยากรณ์ (PRED) และ residual (RESID) เป็นตัวแปรชั่วคราว
(temporary variable, tempvar) เมื่อจะเรียกใช้ให้ใส่ * หน้า
tempvar เสมอ (ดูตัวอย่าง 8)

/ SCATTERPLOT = (varname, varname) SIZE(plotsize)

ใช้พล็อตตัวแปรที่ละคู่ SIZE มี 2 ค่าคือ LARGE กับ SMALL default
คือ SMALL และถ้าจะ plot ตัวแปรชั่วคราวคือ residual หรือ
ค่าพยากรณ์ ให้ใส่ * หน้า varname ถ้าจะพล็อตมากกว่า 1 คู่ ให้สั่ง
คำสั่งการเขียนวงเล็บต่อกัน (ดูตัวอย่าง 9)

/ PARTIALPLOT = varlist SIZE(plotsize)

ใช้สั่งพล็อต residual e_1 กับ e_2 โดยที่ e_1 คือ residual
จากสมการ $Y = f(X\text{'s ยกเว้น } X_j)$ และ e_2 คือ residual
จากสมการ $X_j = f(X\text{'s ยกเว้น } X_j)$ คำสั่งนี้ทำให้เรามองเห็น
ความสัมพันธ์แท้ ๆ ระหว่าง Y กับ X_j ทำให้ปรับรูปสมการได้เหมาะสม
ยิ่งขึ้น ถ้าสั่ง /PARTIALPLOT เฉย ๆ จะทำการพล็อตให้ทุก j

/ CASEWISE ใช้สั่งพิมพ์และพล็อต temporary variable keyword ที่ใช้มี DEFAULTS

OUTLIERS(value) ALL PLOT(tempvars) และ varlist

/ SAVE ใช้เก็บ temporary var รูปแบบคือ tempvar (valid variable)

เพื่อเรียกใช้งานต่อไป โดยเราเรียกใช้ valid var (ดูตัวอย่าง 10)

/ REGWGT ใช้ระบุตัวแปรเพื่อทำ WLS โดยให้ระบุชื่อตัวแปรพร้อมทั้ง weight เช่น

/REGWGT = WGT-1 ในสั่ง /REGWGT หลัง /VARIABLES ก่อน /DEPENDENT

/ WIDTH ใช้ความคมหน้ากว้างของ display ให้กำหนดได้ตามต้องการแต่อย่าให้สั้นเกินไป เพราะผลลัพธ์จะไม่สวย default อยู่ใน SET command

หมายเหตุสำคัญ subcommand /CRITERIA /STATISTICS /ORIGIN /SELECT /MISSING และ /REGWGT ต้องสั่งก่อน subcommand /DEPENDENT

ตัวอย่าง 1

```
COMPUTE LOGBEG = LG10(SALBEG)

REGRESSION VARIABLES = LOGBBG EDLEVEL SEX WORK MINORITY AGE

/DEPENDENT = SALNOW

/METHOD = ENTER SALBEG

/DEPENDENT = LOGBEG

/METHOD = ENTER EDLEVEL TO AGE
```

ตัวอย่าง 2

```
REGRESSION VARIABLES = LOGBEG EDLEVBL SEX WORK MINORITY AGE

/DEPENDENT = LOGBEG

/METHOD = ENTER DELEVEL

/METHOD = FOREWARD SEX TO AGE

/METHOD = BACKWARD
```

ตัวอย่าง 3

```
REGRESSION VARIABLES = . . .

/CRITERIA = FIN FOUT

/DEPENDENT = LOGBEG

/METHOD = STEPWISE
```

ตัวอย่าง 4

```
REGRESSION VARIABLES = . . .  
/STATISTICS = CI  
/DEPENDENT = SALNOW  
/METHOD = ENTER EDLEVEL  
/METHOD = ENTER SEX
```

ตัวอย่าง 5

```
REGRESSION VARIABLES = . . .  
/DEPENDENT = SALNOW  
/METHOD = ENTER SALBEG  
/ORIGIN  
/DEPENDENT = SALBEG  
/METHOD = ENTER EDLEVEL
```

ตัวอย่าง 6

```
REGRESSION SELECT SEX EQ 0  
/VARIABLES = . . .  
/DEPENDENT = LOGBEG  
/METHOD = STEPWISE  
/RESIDUALS = HISTOGRAM
```

ตัวอย่าง 7

```
REGRESSION DESCRIPTIVES = CORR  
/VARIABLES = . . .
```

/DEPENDENT = LOGBEG

/METHOD = ENTER EDLEVEL TO AGE

ตัวอย่าง 8

REGRESSION VARIABLES = ...

/DEPENDENT = SALNOW

/METHOD = ENTER SALBEG

/RESIDUALS

ตัวอย่าง 9

COMPUTE LOGBEG = LG10(SALBEG)

COMPUTE LOGNOW = LG10(SALNOW)

REGRESSION VARIABLES = LOGBEG LOGNOW

/DEPENDENT = LOGNOW

/METHOD = ENTER LOGBEG

/SCATTERPLOT(*SRESID,*PRED) (SALBEG,*PRED)

ตัวอย่าง 10

FILEHANDLE BANK /NAME = 'BANK SPSSXFIL'

GETFILE = BANK

REGRESSION VARIABLES = SALBEG,SALNOW

/DEPENDENT = SALNOW

/METHOD = ENTER SALBEG

/SAVE = SEPREP(SE)

PLOT CUTPOINTS = EVERY(20) / SYMBOL = '.'

/PLOT = SE WITH SALBEG

สรุปคำสั่งของกระบวนการความ FACTOR

subcommand สำหรับ Factor Analysis มีดังนี้

/VARIABLES = ชื่อตัวแปร

เราส่ง /VARIABLES =... ได้ครั้งเดียวในการวิ่ง FACTOR command และ Subcommand นี้ต้องอยู่ก่อน Subcommand อื่น ๆ ยกเว้น /MISSING /WIDTH และ /MATRIX (ดูตัวอย่าง 1)

/MISSING = LISTWISE PAIRWISE MEANSUB INCLUDE

ใช้จัดการกับปัญหา missing obs. ก่อนที่จะส่งคำสั่งเกิดปัญหา corr ระหว่างตัวแปรที่ระบุชื่อใน /VARIABLES ถ้าไม่ส่ง /MISSING หรือส่ง /MISSING เฉย ๆ โดยไม่ระบุ keyword โปรแกรมจะทำ LISTWISE ซึ่งเป็น default เราต้องส่ง(ถ้าต้องการ) /MISSING ก่อนคำสั่งอื่น ยกเว้น /VARIABLES และ /WIDTH (ดูตัวอย่าง 1)

/WIDTH = จำนวน

ใช้กำหนดความกว้างของ display สั่งได้ครั้งเดียวและจะวาง /WIDTH ไว้ที่ใดก็ได้

/ANALYSIS = ชื่อตัวแปรบางส่วน

ใช้สั่งทำ FA สำหรับตัวแปรบางส่วนที่ระบุ เราสามารถทำ multiple subcommand ได้ (ดูตัวอย่าง 2) เรียกว่า analysis block

/EXTRACTION = PC PAF ML ALPHA IMAGE ULS GLS

ใช้สั่งเลือกวิธีจัดกลุ่ม เราสามารถทำ multiple subcommand ได้ และต้องมี /ROTATION รวมด้วยมิเช่นนั้นจะไม่ rotate (ดูตัวอย่าง 3)

/DIAGONAL = value list

ใช้เฉพาะกับวิธี PAF เช่น /DIAGONAL .5 .4 .3 .4 .2

/EXITRACTION = PAF

/CRITERIA = FACTORS(nf) MINEIGEN(eg) ITERATION(ni) ECONVERGE(e1)

RCONVERGE(e2) KAISER NOKAISER DELTA(d) DEFAULT

SPSS กำหนด default ของ subcommand นี้ไว้ดังนี้

nf(=no. of factor) = จำนวน eigen ที่สูงกว่ากำหนดไว้ใน

MINEIGEN eg(eigenvalue) = 1 ni(no. of iteration) = 25

e1 = .001 e2 = .0001 KAISER d=0 (ดูตัวอย่าง 4)

กำหนดแน่นอน ถ้าเราไม่สั่ง /EXTRACTION และ /ROTATION โปรแกรมจะหมุนแกนตามวิธี

VARIMAX ถ้าสั่ง /EXTRACTION แต่ไม่สั่ง /ROTATION โปรแกรมจะไม่ทำการหมุนแกน

แปลว่า VARIMAX เป็น default ของกรณีแรกและ NOROTATE เป็น default ของกรณีหลัง

/ROTATION = VARIMAX EQUAMAX QUARTIMAX OBLIMIN NOROTATE

เราสามารถให้ multiple subcommand นี้ได้ (ดูตัวอย่าง 5)

/PRINT = UNIVARIATE INITIAL CORRELATION SIG DET INV AIC KMO

EXTRACTION REPR ROTATION FSCORE DEFAULTS ALL

ใช้ /PRINT ได้ 1 ครั้งต่อ 1 analysis block

/FORMAT = SORT BLAKL(n) DEFAULTS

ใช้กำหนด display เราใช้ /FORMAT ได้ 1 ครั้งต่อ 1 analysis block

(ดูตัวอย่าง 6)

/PLOT = EIGEN ROTATION (n1 n2)

ใช้พล็อตกราฟของ eigenvalue หรือ factor loading คำสั่ง /PLOT

สั่งได้ 1 ครั้งต่อ 1 analysis block ถ้าจะพล็อตหลายรูปให้วางเส้นคั่น ๆ กัน

หลายวงเส้น คำสั่ง /PLOT ใช้ในกรณี rotated factor ถ้าต้องการพล็อต

unrotated factor ให้สั่ง /ROTATION = NOROTATE ก่อนสั่ง /PLOT

(ดูตัวอย่าง 7)

/SAVE = วิธี (จำนวน factor score ชื่อ)

ใช้คำนวณและเก็บ factor score ไว้ใน active file การสั่ง /SAVE

ต้องกำหนด 3 อย่างคือ ชื่อวิธีหา factor score จำนวน factor score

ที่จะหา และ rootname ของ factor score ที่เราตั้งให้

(factor score สั่งให้หาผ่าน matrix input ไม่ได้)

1) วิธีหา factor score อาจเป็น REG (วิธีนี้เป็น default)

BART(bartlett) หรือ AR (Anderson-Rubin)

2) สิ่งที่ต้องกำหนดในวงเล็บ

2.1 จำนวน factor score มีได้มากค่าแล้วแต่วิธี extraction เรา

อาจสั่งว่า ALL ถ้าต้องการ factor score จากทุกวิธี extraction

2.2 rootname ต้องตั้งไว้ไม่เกิน 7 อักขระ (ดูตัวอย่าง 8)

ตัวอย่าง 1

FACTOR VARIABLES = ...

/MISSING = MEANSUB

/WIDTH = 80

ตัวอย่าง 2

FACTOR VARIABLES = V1 V2 V3 V4 V5 V6 V7

/PRINT = CORRELATION

/ANALYSIS = V1 V2 V3 V4

/ANALYSIS = V1 V3 V4 V5 V6 V7

ตัวอย่าง 3

FACTOR VARIABLES = V1 V2 V3 V4 V5 V6

/EXTRACTION = PC

/EXTRACTION = ML

ตัวอย่าง 4

FACTOR VARIABLES = IQ GPA TESTSCOR

STRESS SAT PSYCHTST

/CRITERIA = FACTOR(2)

/ANALYSIS = ALL

/CRITERIA = DEFAULT

ตัวอย่าง 5

FACTOR VARIABLES = ...

/PLOT = EIGEN ROTATION(1,2)

/ROTATION = NOROTATE

จะทำการพล็อต factor loading ของ F1 และ F2 ขณะก่อนหมุนแกน

ตัวอย่าง 6

FACTOR VARIABLES = POPSTABL TO MENTALIL

/FORMAT = SORT BLANK(.5)

ตัวอย่าง 7

FACTOR VARIABLES = ...

/PLOT = EIGEN ROTATION(1 2)(1 3)(2,3)

ตัวอย่าง 8

```
FACTOR VARIABLES = ABDEWCT TO ABSINGLE
```

```
/MISSING = MEANSUB
```

```
/CRITERIA = FACTOR(2)
```

```
/EXTRACTION = ULS
```

```
/ROTATION = VARIMAX
```

```
/SAVE = AR (ALL FSULS)
```

คำนวณ factor score บัญชีชื่อ FSULS1 กับ FSULS2 โดยวิธี AR แล้วเก็บผลไว้ใน active file เราอาจสั่งในลักษณะ multiple SAVE ดังนี้

```
FACTOR VARIABLES = ABDEFECT TO ABSINGLE
```

```
/MISSING = MEANSUB
```

```
/EXTRACTION = ULS
```

```
/ROTATION = VARIMAX
```

```
/SAVE AR(ALL FSULS)
```

```
/SAVE BART(ALL BFAC)
```

สรุปคำสั่งของกระบวนการความ CLUSTER

สิ่งที่ต้องกำหนดสำหรับ command นี้มีเพียง variable list เท่านั้น ซึ่งต้องระบุก่อนเข้า subcommand อื่น ๆ ทั้งหมด subcommand ทั้งปวงที่จะกล่าวถึงต่อไปนี้เป็นสิ่งที่เราอาจเลือกสั่งหรือไม่ก็ได้ ข้อมูลที่จะนำเข้าสู่ command นี้จะได้รับการปรับเป็นคะแนนมาตรฐานเสียก่อน (ดูตัวอย่าง 1)

```
/METHOD = BAVERAGE WAVERAGE SINGLE COMPLETE CENTROID MEDIAN WARD
```

หมายเหตุ CENTROID MEDIAN WARD ใช้กับ /MEASURE = SEUCLID เท่านั้น

(ดูตัวอย่าง 2)

BAVERAGE เป็น default เมื่อไม่สั่ง /METHOD จะทำด้วยวิธีที่เป็น default
/MEASURE = SEUCLID EUCLID COSINE BLOCK CHEBYCHEV POWER (p,r) DEFAULT
ถ้าไม่สั่ง /MEASURE จะทำการวิเคราะห์จาก similarity measure วิธี
square euclidian distance คือ SEUCLID เรากำหนด distant
measure ได้เพียงอย่างเดียว

/PRINT = SCHEDULE CLUSTER(min max) DISTANCE NONE

ถ้าไม่สั่ง /PRINT โปรแกรมจะพิมพ์ default คือ SCHEDULE ขอแนะนำให้ใช้
keyword CLUSTER (min,max) เพราะแปลผลง่ายกว่าวิธีอื่น

/PLOT = VICICLE (min,max,inc) HICICLE(min,max,inc) DENDROGRAM,NONE

VICICLE เป็น default min max inc (คือจำนวน cluster ต่ำสุด สูงสุด
และ increment) ต้องเป็นจำนวนเต็ม

/ID = default

default จะใช้ case number เป็น ID เว้นแต่กำหนดเป็นอย่างอื่นด้วย /ID
เช่น /ID = BEER

MISSING = LISTWISE INCLUDE

default คือ LISTWISE

ตัวอย่าง 1

CLUSTER ZV1 ZV2 ZV3 ZV4

ตัวอย่าง 2

CLUSTER ZV1 ZV2 ZV3 ZV4

/METHOD = SINGLE COMPLETE

รายละเอียดคำสั่งของกระบวนการความ REGRESSION

REGRESSION

```

REGRESSION {/WIDTH = {132
n t}}

{/SELECT = {(ALL)
{varname relation value}}

{/MISSING = {LISTWISE**
PAIRWISE
MEANSUBSTITUTION
INCLUDE}}

{/DESCRIPTIVE = [DEFAULTS] {MEAN} {STDDEV} {CORR}
[VARIANCE] {XPROD} [SIG] [N] {BADCORR}
[COV] [NONE]]

/VARIABLES = {varlist}

{/CRITERIA = {DEFAULTS**} [PIN({0.05
value})] [POUT({0.1
value})]

[TOLERANCE({0.01
value})]

[MAKSTEPS({2v
n})] [FIN({3.84
value})] [FOUT({2.71
value})]

{/STATISTICS = {DEFAULTS**} [R] [COEFF] [ANOVA] [ZPP]
[LABEL] [CHA] [C I] [F] [BCOV] [SES]
[HISTORY] [XTX] [COND] [END] [ALL]]

{/ NOORIGIN}
{ORIGIN}}

/DEPENDENT = varlist

{STEPWISE [= varlist]
FORWARD [= varlist]
BACKWARD [= varlist]} [I] [/...]
{ENTER [= varlist]
REMOVE = varlist}

{/RESIDUALS = {DEFAULTS} {DURBIN}

[HISTOGRAM {ZRESID
tempvar}] [OUTLIERS {ZRESID
tempvar}]

[NORMPROB {ZRESID
tempvar}] [SIZE {LARGE
SMALL}]]

{/CASEWISE = [DEFAULTS] [OUTLIERS {3
value}] [PLOT {ZRESID
tempvar.}]

[DEPENDENT] [PRED] [RESID] [tempvarlist] [ALL]]

{/SCATTERPLOT = [SIZE {SMALL
LARGE}] (varname, varname) ]

```

****Default if the subcommand is omitted.**

Temporary residual variables are:

PRED ADJPRED Z P R E D SEPRED RESID ZRESID DRESID
SRESID SDRESID MAHAL C O O K

6.56 RUNNING PROCEDURE REGRESSION

The REGRESSION procedure provides five equation-building methods: forward selection, backward elimination, stepwise selection, forced entry, and forced removal. The subcommands for residual analysis help detect influential data points, outliers, and violations of the regression model assumptions.

6.57 Building the Equation

To build a simple regression model, you must specify three required subcommands: a VARIABLES subcommand that names the variables to be analyzed, a DEPENDENT subcommand that indicates the dependent variable, and a METHOD subcommand that names the method to be used. For example, to build the simple bivariate model of beginning salary and current salary discussed earlier in the chapter, specify

```
REGRESSION VARIABLES=SALBEG SALNOW  
/DEPENDENT=SALNOW  
/METHOD=ENTER SALBEG.
```

The beginning (SALBEG) and current (SALNOW) salaries are named, with the latter specified as the dependent variable. The ENTER keyword enters beginning salary into the equation. The output produced from this command is shown in Figures 6.3b, 6.10, and 6.11a.

6.58 VARIABLES Subcommand

The optional VARIABLES subcommand lists all variables to be used in the regression analysis. The order of variables on the VARIABLES subcommand determines the order of variables in the correlation matrix. The keyword TO can be used on the VARIABLES subcommand to imply consecutive variables on the active system file. On subsequent DEPENDENT and METHOD subcommands, the keyword TO refers to the order of variables on the VARIABLES subcommand.

The VARIABLES subcommand is followed by a variable list or either one of the following keywords:

ALL *Include all user-defined variables in the active system file.*
(COLLECT) *Include all variables named on the DEPENDENT and METHOD subcommands. This is the default if the VARIABLES subcommand is not included.*

If you do not include a VARIABLES subcommand or you specify the keyword (COLLECT), the METHOD subcommand(s) must include a variable list. If used, the VARIABLES subcommand must precede the first DEPENDENT and METHOD subcommands, as in:

```
REGRESSION VARIABLES=SALBEG SALNOW LOGBEG  
EDLEVEL SEX WORK MINORITY AGE  
/DEPENDENT=LOGBEG  
/METHOD=ENTER EDLEVEL TO AGE.
```

6.59
DEPENDENT Subcommand

The **DEPENDENT** subcommand indicates the dependent **variable** for the **regression** analysis. The **DEPENDENT** subcommand is followed by a variable name or variable list. If you specify more than one variable on the **DEPENDENT** s&command, **SPSS** produces a separate equation for each dependent variable **specified**.

You can specify more than one **analysis** with multiple **DEPENDENT** and **METHOD** subcommands. For example, to run both a bivariate and **multivariate** analysis in the same **REGRESSION** procedure, specify

```
REGRESSION VARIABLES=SALBEG SALNOW LOGBEG
EDLEVEL SEX WORK MINORITY AGE
/DEPENDENT=SALNOW
/METHOD=ENTER SALBEG
/DEPENDENT=LOGBEG
/METHOD=ENTER EDLEVEL TO AGE.
```

The **first** **DEPENDENT** subcommand **defines** a single equation with **SALNOW** as the dependent variable, and the **METHOD** s&command enters **SALBEG** into the equation. The second **DEPENDENT** subcommand defines another equation, with **LOGBEG** as the dependent variable. The associated **METHOD** subcommand enters variables **EDLEVEL** to **AGE** into the equation. The **TO** convention for naming consecutive variables used in the second **METHOD** subcommand refers to the order in which the variables are named on the **VARIABLES** subcommand, not their order on the active system file. See Figures 6.36a and 6.36b for the **output** from the second equation.

If you specify more than one variable on the **DEPENDENT** subcommand, **SPSS** produces a separate equation for each dependent variable specified.

6.60
METHOD Subcommand

At least one **METHOD** subcommand must immediately follow each **DEPENDENT** subcommand, specifying the method to be used in developing the regression equation. The available methods are

- FORWARD** (varlist) *Forward variable selection.* Variables are entered one at a time based on entry **criteria** (Section 6.46).
- BACKWARD** (varlist) *Backward variable elimination.* All variables are entered and then **removed one at a time** based on **removal criteria** (Section 6.47).
- STEPWISE** (varlist) *Stepwise variable entry and removal.* Variables are **examined at each step** for entry or removal (Section 6.48).
- ENTER** (varlist) *Forced entry.* The variables named are entered in a single step. The default **variable list** is all independent **variables**.
- REMOVE** (varlist) *Forced removal.* The variables named are removed in a **single step**. **REMOVE** must have an **accompanying variable list**.
- TEST** (varlist) *Test indicated subsets of independent variables.* **TEST** offers an **easy way to test** a variety of models using R^2 change and its test of **significance** as the criterion for the "best" model. **TEST** must have an **accompanying variable list**.

A variable list is required with the **REMOVE** and **TEST** keywords and is optional for the other **METHOD** keywords. The default variable list for methods **FORWARD**, **BACKWARD**, **STEPWISE**, and **ENTER** includes all variables named on the **VARIABLES** subcommand that are not named on the **preceding DEPENDENT** subcommand. For example, to request the backward-elimination method discussed in Section 6.47, **specify**


```
REGRESSION VARIABLES=LOGBEG EDLEVEL SEX WORK MINORITY AGE
/DEPENDENT=LOGBEG
/METHOD=BACKWARD.
```

The keyword METHOD is optional and may be omitted. For example, the command

```
REGRESSION VARIABLES=LOGBEG EDLEVEL SEX WORK MINORITY AGE
/DEPENDENT=LOGBEG
/BACKWARD.
```

produces the same results as the previous example.

You can specify multiple METHOD subcommands. For example, you might want to force one variable into the equation first and then enter the remaining variables in a forward-selection fashion, as in

```
REGRESSION VARIABLES=LOGBEG EDLEVEL SEX WORK MINORITY AGE
/DEPENDENT=LOGBEG
/METHOD=ENTER EDLEVEL
/METHOD=FORWARD SEX TO AGE.
```

6.61
STATISTICS
Subcommand

By default, REGRESSION displays the four sets of statistics described for keywords R, ANOVA, COEFF, and OUTS below. These statistics are shown in Figures 6.3b, 6.10, and 6.11 a for the bivariate equation, and in Figures 6.36a and 6.36b for the multivariate equation. You can specify exactly which statistics you want displayed by any of the following keywords on the STATISTICS subcommand.

- DEFAULTS** R, ANOVA, COEFF, and OUTS. These statistics are displayed when the STATISTICS subcommand is omitted or if no keywords are specified on the subcommand. If you specify statistics keywords on a STATISTICS subcommand, the default statistics will not appear unless you specify them explicitly, either individually or with the DEFAULTS keyword.
- ALL** All statistics except F, LINE, and END.
- R** Multiple R. Displays multiple R, R', adjusted R², and the standard error. (See Figure 6.10.)
- ANOVA** Analysis @variance table. Displays degrees of freedom, sums of squares, mean squares, F value for multiple R, and the observed significance level of F. (See Figure 6.11 a.)
- CH4** Displays change in R² between steps, F value for change in R², and significance of F. (See Figure 6.41.)
- BCOV** Variance-covariance matrix. Displays a matrix with covariances above the diagonal, correlations below the diagonal, and variances on the diagonal.
- ATX** Sweep matrix.
- COLLIN** Collinearity diagnostics. Includes the variance inflation factor (VIF), the eigenvalues of the scaled and uncentered cross-products matrix, condition indices, and variance-decomposition proportions (Belsley et al., 1980).
- SELECTION** Aids to selecting set of regressors. Includes Akaike information criterion (AIC), Amemiya's prediction criterion (PC), Mallows's conditional mean squared error of prediction criterion (Cp), and Schwarz Bayesian criterion (SBC) (Judge et al., 1985).
- COEFF** Statistics for variables in the equation. Displays regression coefficient B, standard error of B, standardized coefficient Beta, t value for B, and two-tailed significance level of t.

OUTS	<i>Statistics for variables nor in the equation</i> that have been named on the VARIABLES subcommand . Statistics are Beta if the variable were entered, <i>t</i> value, for Beta, significance level off, partial correlation with the dependent variable controlling for variables in the equation, and minimum tolerance. (See Figure 6.42.)
ZPP	<i>Zero-order, part, and partial correlation.</i> (See Figure 6.39.)
CI	Confidence intervals. Displays the 95% confidence interval for the unstandardized regression coefficient . (See Figure 6.8.)
SES	<i>Approximate standard error of the standardized regression coefficients.</i> (Meyer and Younger, 1976)
TOL	Tolerance. Displays tolerance and VIF for variables in the equation and for variables not in the equation, the tolerance a variable would have if it were the only variable entered next. (See Figure 6.5 1.)
F	F value for B and significance of F. Displayed instead of <i>t</i> for COEFF and OUTS . (See, for example , Figures 6.464 6.46c, 6.47a, and 6.47b.)
LINE	Summary <i>line for each step in step methods</i> . Displays a single summary line for each step in BACKWARD , FORWARD , or STEPWISE methods and the default or requested statistics at the end of each method block (BACKWARD , FORWARD , STEPWISE , ENTER , REMOVE , or TEST).
HIS-TORY	Step history. Displays a summary report with a summary line for each method (ENTER , REMOVE , or TEST , if the equation changes) or step if the method entails steps (FORWARD , BACKWARD , or STEPWISE). If history is the only statistic requested, COEFF is displayed for the final equation. (See Figures 6.43 and 6.46a.)
END	<i>One summary line per step or method block.</i> Displays a summary line per step for BACKWARD , FORWARD , or STEPWISE , and one summary line per block for ENTER , REMOVE , or TEST , if the equation changes.

The **STATISTICS** subcommand must appear before the **DEPENDENT subcommand** that initiates the equation and remains in effect until overridden by another **STATISTICS** subcommand. For example, to produce the output in Figure 6.8, **specify**

```
REGRESSION VARIABLES=SALBEC SALNOW
/STATISTICS=CI
/DEPENDENT=SALNOW
/METHOD=ENTER SALBEG.
```

To produce the output for the **multivariate** example shown in Figure 6.41, **specify**

```
REGRESSION VARIABLES=LOGBEG EDLEVEL SEX WORK MINORITY AGE
/STATISTICS=CI
/DEPENDENT=LOGBEG
/METHOD=ENTER EDLEVEL
/METHOD=ENTER SEX.
```

6.62 CRITERIA Subcommand

You can control the statistical criteria by which **REGRESSION** chooses variables for entry into or removal from an equation with the **CRITERIA subcommand**. Place the **CRITERIA subcommand** after the **VARIABLES** subcommand and before the **DEPENDENT subcommand**. A **CRITERIA** subcommand affects **any** subsequent **DEPENDENT** and **METHOD** subcommands and remains in effect until overridden with another **CRITERIA** subcommand.

The **CRITERIA** keywords are

DEFAULTS *PIN(0.05), POUT(0.10), and TOLERANCE(0.0001)*. These are the defaults if no **CRITERIA subcommand** is specified. **If** criteria have been changed, **DEFAULTS** restores the default values.

PIN(value)	<i>Probability of F-to-enter.</i> Use to override the default value of 0.05.
POUT(value)	<i>Probability of F-to-remove.</i> Use to override the default value of 0.10.
FIN(value)	<i>F-to-enter</i> The default value is 3.84. FIN and PIN are mutually exclusive .
FOUT(value)	<i>F-to-remove.</i> The default value is 2.71. FOUT overrides the default POUT criteria. If both FOUT and POUT are specified on the same CRITERIA subcommand, only the last one specified will be in effect .
TOLERANCE(value)	<i>Tolerance.</i> The default value is 0.0001 . All variables must pass both tolerance and minimum tolerance tests before entering the equation. The minimum tolerance is the smallest tolerance for that variable or any other variable in the equation if the variable is entered.
MAXSTEPS(n)	<i>Maximum</i> number of <i>steps</i> . For the STEPWISE method, the default is twice the number of independent variables. For the FORWARD and BACKWARD methods, the default maximum is the number of variables meeting the PIN and POUT or FIN and FOUT criteria. The MAXSTEPS value applies to the total model. The default value for the total model is the sum of the maximum number of steps over each method in the model.

For example, to change **stepwise** entry and removal criteria to FIN and **FOUT** and use their default values of 3.84 and 2.71, respectively, specify

```
REGRESSION VARIABLES=LOGBEG EDLEVEL SEX WORK MINORITY ACE
/CRITERIA=FIN,FOUT
/DEPENDENT=LOGBEG
/METHOD=STEPWISE.
```

6.63 ORIGIN Subcommand

The regression model contains a constant term. You **can** use the **ORIGIN subcommand** to suppress this term and obtain regression through the origin. The **NOORIGIN subcommand**, which is the default, requests that equations include a constant term.

Place the **ORIGIN** or **NOORIGIN** subcommand **between** the **VARIABLES** subcommand and the **DEPENDENT** subcommand for the equation. For example,

```
REGRESSION VARIABLES=SALBEG SALNOW, EDLEVEL
/DEPENDENT=SALNOW
/METHOD=ENTER SALBEG
/ORIGIN
/DEPENDENT=SALBEG
/METHOD=ENTER EDLEVEL.
```

requests two equations, the first with a constant term (the default) **and** the second with regression through the origin.

There are no specifications for the **ORIGIN** and **NOORIGIN** subcommands. Once specified, the **ORIGIN subcommand** remains in effect until **NOORIGIN** is requested.

6.64 SELECT Subcommand

Use the **SELECT** subcommand to select a subset of cases for computing the regression equation. **Only selected** cases contribute to the correlation coefficients and to the regression equation. Residuals and predicted values are calculated and reported separately for both selected and unselected cases. The **SELECT** subcommand can precede or immediately follow the **VARIABLES** subcommand and is in effect for the entire **REGRESSION command**. The form of the **SELECT** subcommand is

/SELECT= **varname** relation value

The relation can be EQ, NE, **LT**, LE, GT, or GE.

For example, to generate separate residuals histograms for males and females based on the equation developed for males alone (**SEX=0**) as shown in Figure 6.55, specify

```
REGRESSION SELECT SEX EQ 0
/VARIABLES=LOGBEG EDLEVEL SEX WORK MINORITY AGE
/DEPENDENT=LOGBEG
/METHOD=STEPWISE
/RESIDUALS=HISTOGRAM.
```

6.65 MISSING Subcommand

Use the MISSING subcommand to specify the treatment of cases with missing values. If the MISSING subcommand is omitted, a **case** with user- or **system-missing values** for any variable named on the VARIABLES subcommand is excluded from the computation of the correlation matrix on which all analyses are based. The MISSING subcommand can precede or immediately follow the VARIABLES subcommand and is in effect for the entire REGRESSION command.

The available keywords are

- LISTWISE** *Delete cases with missing values listwise.* Only cases with valid values for **all** variables listed on the VARIABLES subcommand are included in analyses. If INCLUDE is also specified, only cases with system-missing values are deleted listwise. **LISTWISE** is the default.
- PAIRWISE** *Delete cases with missing values pairwise.* Cases with complete data on the pair of variables being correlated are used to **compute** the correlation coefficient. If INCLUDE is also specified, only cases with system-missing values are deleted pairwise.
- MEANSUBSTITUTION** *Replace missing values with the variable mean.* All cases are used for **computations**, with the mean of a variable substituted for **missing observations**. If INCLUDE is also specified, user-missing values are included in the computation of the means and only system-missing values are substituted.
- INCLUDE** *include all cases with user-missing values.* Only cases with **system-missing values** are excluded.

If you specify any combination of LISTWISE, PAIRWISE, and MEANSUBSTITUTION on the same MISSING subcommand, only the last one specified will be in effect. If INCLUDE is also specified, it will also be in effect.

6.66 DESCRIPTIVES Subcommand

You can request a variety of descriptive statistics with the DESCRIPTIVES subcommand. These statistics are displayed for **all** variables specified on the VARIABLES subcommand, regardless of which variables you specify for computations. **Descriptive** statistics are based on **all** valid cases for each variable if you have specified PAIRWISE or MEANSUB on the MISSING subcommand. Otherwise, only cases that are included in the computation of the correlation matrix are used. If you specify the DESCRIPTIVES subcommand without any keywords, the statistics listed for **keyword DEFAULTS** are displayed. If you name any statistics on DESCRIPTIVES, only those explicitly requested are displayed.

The following descriptive statistics are available:

DEFAULTS *MEAN, STDDEV, and CORR* This is the default if **DESCRIPTIVES** is specified without any keywords.

MEAN *Variable means.*

STDDEV *Variable standard deviations.*

VARIANCE *Variable variances.*

CORR *Correlation matrix.*

SIG *One-tailed significance levels for the correlation coefficients.*

BADCORR *Correlation matrix only if some coefficients cannot be computed.*

COV *Covariance matrix.*

WROD *Cross-product deviations from the mean*

N *Number of cases used to compute the correlation coefficients.*

A U *All descriptive statistics.*

For example, to produce the correlation matrix shown in Figure 6.34, specify

```
REGRESSION DESCRIPTIVES=CORR
/VARIABLES=LOGBEG EDLEVEL SEX WORK MINORITY ACE
/DEPENDENT=LOGBEG
/METHOD=ENTER EDLEVEL TO AGE.
```

6.67 Analyzing Residuals

Once you have built an equation, REGRESSION can calculate a variety of temporary variables containing several types of residuals, predicted values, and related measures. You can use these variables to detect outliers and influential data points and to examine the regression assumptions described in Sections 6.17 through 6.22.

The following temporary variables are available for the analysis of residuals.

PRED *Unstandardized predicted values.* (See Section 6.13.)

ZPRED *Standardized predicted values.* (See Section 6.13.)

SEPPRED *Standard errors of the predicted values.* (See Section 6.14.)

RESID *Unstandardized residuals.* (See Section 6.18.)

ZRESID *Standardized residuals.* (See Section 6.18.)

SRESID *Studentized residuals.* (See Section 6.18.)

MAHAL *Mahalanobis' distance.* (See Section 6.24.)

ADJPRED *Adjusted predicted values.* (See Section 6.25.)

DRESID *Deleted residuals.* (See Section 6.25.)

SDRESID *Studentized deleted residuals.* (See Section 6.25.)

COOK *Cook's distances.* (See Section 6.25.)

LEVER *Leverage values.* (See Section 6.49.)

DFBETA *DFBETA.* The change in the regression coefficient that results from the deletion of the *i*th case. A DFBETA value is computed for each case for each regression coefficient generated in a model.

SDBETA *Standardized DFBETA.* An SDBETA value is computed for each case for each regression coefficient generated in a model. (See Belsley et al., 1980.)

DFFIT *DFFIT.* DFFIT is the change in the predicted value when the *i*th case is deleted. (See Belsley et al., 1980.)

SDFIT *Standardized DFFIT.* (See Belsley et al., 1980.)

COVRATIO *COVRATIO.* Ratio of the determinant of the covariance matrix with the *i*th case deleted to the determinant of the covariance matrix with all cases included. (See Belsley et al., 1980.)

MCIN *Lower and upper bounds for the prediction interval of the mean predicted response. A lower bound LMCIN and an upper bound UMCIN are generated. The default confidence interval is 95%. The interval may be reset with the CIN subcommand. (See Dillon & Goldstein, 1984.)*

ICIN *Lower and upper bounds/or the prediction interval for a single observation. (See Dillon & Goldstem, 1978.) A lowerbound LICIN and an upperbound UICIN are generated. The default confidence interval is 95%. The interval may be reset with the CIN subcommand.*

Residuals **analysis** is specified with four subcommands: RESIDUALS, CASEWISE, PARTIALPLOT, and SCATTERPLOT. You can specify these subcommands in any order. but you cannot specify more than one of each per equation.

and they must immediately follow the last METHOD subcommand that completes an equation. The residuals subcommands affect only the equation they follow. Requesting any residuals analysis always produces descriptive statistics on at least four of the temporary variables (PRED, ZPRED, RESID, and ZRESID).

All variables are standardized before plotting. If an unstandardized version of a variable is requested. the standardized version is plotted.

6.66 RESIDUALS Subcommand

Use the RESIDUALS subcommand to obtain the statistics and plots listed below. Specifying the RESIDUALS subcommand without any specifications produces the display described for keyword DEFAULTS. If any keywords are specified on RESIDUALS. only the displays for those keywords are produced.

DEFAULTS *HISTOGRAM(ZRESID), NORMPROB(ZRESID), OUTLIERS plots(ZRESID), SIZE(SMALL), and DURBIN.* These plots are produced if RESIDUALS is specified without any specifications.

HISTOGRAM(tempvars) *Histogram of standardized temporary variables named.* The default temporary variable is ZRESID. Other variables that can be plotted are PRED, RESID, ZPRED, DRESID, ADJPRED, SRESID, and SDRESID. (See Figure 6.22a.)

NORMPROB(tempvars) *Normal probability (P-P) plot of standardized values.* The default variable is ZRESID. Other variables that can be plotted are PRED, RESID, ZPRED, and DRESID. (See Figure 6.22b.)

SIZE(plotsize) *Plot sizes.* The plot size can be specified as SMALL or LARGE. The default is LARGE if the display width is at least 120 and the page length is at least 55.

OUTLIERS(tempvars) *The ten most extreme values for the temporary variables named.* The default temporary variable is ZRESID. Other variables can be RESID, DRESID, SRESID, SDRESID MAHAL, and COOK. (See Figure 6.24.)

DURBIN *Durbin-Watson test statistic. (See Section 6.2 1.)*

ID(varname) *Identification labels for casewise and outlier plots. Cases are labeled with values of the variable named after the ID keyword. By default, the plots are labeled with the sequential case number. ID also labels the CASEWISE list of cases. (See Figures 6.23 and 6.24.)*

POOLED *Pooled plots and statistics when the SELECT subcommand is in effect. All cases in the active tile are used. The default is separate reporting of residuals statistics and plots for selected and unselected cases.*

For example, to produce the output shown in Figures 6.22a, 6.22b, and 6.24, specify

```
REGRESSION VARIABLES=SALBEG SALNOW
/DEPENDENT=SALNOW
/METHOD=ENTER SALBEG
/RESIDUALS=HISTOGRAM(SRESID) NORMPROB
OUTLIERS(MAHAL) ID(SEXRACE) SIZE (SMALL).
```

6.69

CASEWISE Subcommand

You can display a **casewise** plot of one of the temporary variables accompanied by a listing of the values of the **dependent** and the **temporary variables**. The plot can be requested for all cases or limited to **outliers**. Specifying the **CASEWISE** subcommand without keywords produces the output listed for **DEFAULTS**.

The following may be specified on the **CASEWISE** subcommand.

DEFAULTS *OUTLIERS(3), PLOT(ZRESID), DEPENDENT, PRED, and RESID.* This is the default if **CASEWISE** is specified without any keywords.

OUTLIERS(value) *Limit plot to outliers greater than or equal to the standardized absolute value of the plotted variable.* The default value is 3. (See Figure 6.23.)

A U *Include all cases in the casewise plot.* Produces a plot of all cases, including outliers. The keyword **OUTLIERS** is ignored when **ALL** is specified.

PLOT(tempvar) *Plot the standardized values of the temporary variable named.* The default variable is **ZRESID**. The other variables that can be plotted are **RESID**, **DRESID**, **SRESID**, and **SDRESID**. (See Figure 6.23.)

varlist *List values of the DEPENDENT and temporary variables named.* Any temporary variable, including **LEVER**, can be listed. The defaults are **DEPENDENT** (the dependent variable), **PRED**, and **RESID**. (See Figures 6.16 and 6.23.)

For example, to produce the **casewise** plot shown in Figure 6.16, specify

```
REGRESSION VARIABLES=SALBEG SALNOW
/DEPENDENT=SALNOW
/METHOD=ENTER SALBEG
/RESIDUALS=ID(SEXRACE)
/CASEWISE=ALL DEPENDENT PRED RESID SEPRED.
```

To plot **outliers** whose absolute values are **equal** to or greater than 3 based on **ZRESID**, you need only specify the **CASEWISE** subcommand. To base the plot on **Studentized** residuals and label it with an **ID** variable, as shown in Figure 6.23, specify

```
REGRESSION VARIABLES=SALBEG SALNOW
/DEPENDENT=SALNOW
/METHOD=ENTER SALBEG
/RESIDUALS=ID(SEXRACE)
/CASEWISE=PLOT(SRESID).
```

If you request more variables than will fit on the page **width** set either with the **SET WIDTH** command or the **WIDTH** subcommand in **REGRESSION**, your output will be truncated (see Section 6.75).

6.70

SCATTERPLOT Subcommand

Use the **SCATTERPLOT** subcommand to generate scatterplots for the variables in the equation. You must name at least one pair of variables on the **SCATTERPLOT** subcommand. Optionally, you can specify the **SIZE** keyword to control the size of the plots. All scatterplots are standardized.

The specifications for **SCATTERPLOT** are

(varname,varname) *The pair of variables to be plotted.* Available variables are PRED, RESID, ZPRED, ZRESID, DRESID, ADJPRED, SRESID, SDRESID, and any variable named on the **VARIABLES** subcommand. Temporary variables should be preceded by an asterisk **on** this subcommand.

SIZE(plotsize) *Plot sizes.* Plot size can be **SMALL** or **LARGE**. The default is **SMALL**.

The first variable named inside the parentheses is plotted on the vertical (**Y**) axis, and the second is plotted on the horizontal (**X**) axis. For example, to generate the **scatterplot** shown in **Figure 6.20**, specify

```
REGRESSION VARIABLES=SALBEG SALNOW
/DEPENDENT=SALNOW
/METHOD=ENTER SALBEG
/SCATTERPLOT=( *SRESID, *PRED).
```

To produce a **scatterplot** for **SRESID** and **PRED** based on the logarithmic transformation of both the dependent and **independent** variables, as shown in **Figure 6.30a**, use the **SCATTERPLOT** sub-command above along with the following transformation commands:

```
COMPUTE LOGBEG=LG10(SALBEG).
CCYPUTE LOGNOW=LG10(SALNOW).
REGRESSION VARIABLES=LOGBEG, LOGNOW
/DEPENDENT=LOGNOW
/METHOD=ENTER LOGBEG
/SCATTERPLOT=( *SRESID, *PRED).
```

To produce more than one scatterplot, simply add pairs of variable names in parentheses, as in

```
/SCATTERPLOT=( *SRESID, *PRED) (SALBEG, *PRED)
```

6.71 PARTIALPLOT Subcommand

Use the **PARTIALPLOT** subcommand to generate partial residual plots. Partial residual plots are scatterplots of the residuals of the dependent variable and an independent variable when both variables are **regressed** on the rest of the independent variables.

If no variable list is given **on** the **PARTIALPLOT** subcommand, a partial residual plot is produced **for every** independent variable in the equation. Plots are displayed in descending order of the standard error **of B**. All plots are standardized.

The specifications on the **PARTIALPLOT** subcommand are

varlist *Independent variables to be used in partial residual plot.* At least two independent variables must be in the equation for a partial residual plot to be produced. You can **specify the** keyword **ALL** to obtain the default plots for every independent **variable** in the equation.

SIZE(plotsize) *Plot sizes.* The plot size can be specified as **SMALL** or **LARGE**. The default plot size is **SMALL**.

For example, the following commands produced **Figure 6.49a**:

```
COMPUTE LOGBEG=LG10(SALBEG).
REGRESSION VARIABLES=LOGBEG SEX MINORITY EDLEVEL WORN
/DEPENDENT=LOGBEG /METHOD=STEPWISE
/PARTIALPLOT=EDLEVEL.
```

6.72 SAVE Subcommand

Use the **SAVE** subcommand to save any or all of the temporary variables described in Section 6.67. The format is the name of the temporary variable followed by a **valid** variable name in parentheses, as in


```

GET FILE='BANK.SYS'
REGRESSION VARIABLES=SALBEG, SALNOW
/DEPENDENT=SALNOW
/METHOD=ENTER SALBEG
/SAVE=SE PRED(SE).
PLOT CUTPOINTS=EVERY(20) /SYMBOLS='*'
/PLOT=SE WITH SALBEG.

```

This example saves the standard **errors** of the predicted values with variable name SE. Then the PLOT procedure is used to plot the standard errors against the values of the independent variable **SALBEG**. Figure 6.14a shows the plot.

If you don't specify a new variable name, SPSS generates a new variable name by default.

If you specify **DFBETA** or **SDBETA**, the number of new variables saved is equal to the total number of variables in the equation, including the constant. For example, the command

```

REGRESSION DEPENDENT=SALBEG
/METHOD=ENTER AGE SEX
/SAVE=DFBETA(DFBVAR)

```

will create and **save** three new variables with the names **DFBVARO**, **DFBVARI**, and **DFBVAR2**.

You can use the keyword **FITS** to automatically **save** the temporary variables **DFFIT**, **SDFIT**, **DFBETA**, **SDBETA**, and **COVRATIO**, as in:

```
/SAVE=FITS.
```

If you specify **FITS**, you cannot specify **new** variable names. SPSS **automatically** generates new variable names.

6.73 READ and WRITE Subcommands

Procedure **REGRESSION** can read and write matrix materials, which can be processed more quickly than cases. Use the **WRITE** subcommand to write **matrix materials** to a file. You can write default matrix materials or specify the materials you want to write, including variable means, standard deviations, variances, a correlation or **covariance** matrix, and the number of cases used to compute the correlations or **covariances**. You can then **use** the **READ** subcommand to read the **matrix materials** into **REGRESSION** for additional analysis.

The **READ** subcommand can also read matrix materials written by other procedures, such as **CORRELATION**, or entered as data in free or fixed format. See Command Reference: Regression-Matrix **Materials** for complete instructions on using matrix materials with **REGRESSION**.

6.74 REGWGT Subcommand

The **REGWGT** subcommand specifies a variable for estimating weighted **least-squares** models. The only specification on **REGWGT** is the name of the single variable containing the weights, as in

```

REGRESSION VARIABLES=IQ TO ACHIEVE /REGWGT=WGT1
/DEPENDENT=VARY /METHOD=ENTER /SAVE=PRED(P) RESID(R).

```

REGWGT remains in effect for **all** analyses specified on the **REGRESSION** command. If you specify more than one **REGWGT** subcommand, only the last one specified will be in effect.

6.75 WIDTH Subcommand

You can use the **WIDTH subcommand** to control the width of the display produced by the **REGRESSION procedure**. The default is the width specified on the **SET** command. The **WIDTH subcommand** in **REGRESSION** overrides the width **specified** on **SET**.

You can use the **WIDTH subcommand** to **change** the appearance of your output. For example, in **Figure 6.41** statistics for **variables** in the equation and variables not in the equation **are** displayed side by side. In **Figure 6.46b**, the **command**

```
REGRESSION WIDTH=80
/VARIABLES=LOGBEG EDLEVEL SEX WORK MINORITY AGE
/STATISTICS=F /DEPENDENT=LOGBEG /METHOD=FORWARD.
```

displays the statistics for **variables not** in the equation below the statistics for variables in the equation.

A smaller page width limits the **number** of statistics that can be displayed in a summary line and **may** also **cause casewise** output to be truncated (see Section **6.69**). Specifying a smaller page width may **also** reduce the size of scatter and normal-probability **plots** in the residuals output.

6.76 Annotated Example

To **produce** the **stepwise** variable selection example discussed in Section **6.48**, **specify**

```
DATA LIST FILE='EMPLOYEE.DAT'
/IO 1-4 SALBEG 6-10 SEX 12 TIME 14-15
ACE 17-20 (2) SALNOW 22-26 EDLEVEL 28-29
WORK 31-34 (2) JOBCAT MINORITY 35-37.
VAR LABELS ID 'EMPLOYEE CODE'
SALBEG 'BEGINNING SALARY'
SEX 'SEX OF EMPLOYEE' /
TIME 'JOB SENIORITY' /
AGE 'AGE OF EMPLOYEE' /
SALNOW 'CURRENT SALARY'
EDLEVEL 'EDUCATIONAL LEVEL' /
WORK 'WORK EXPERIENCE' /
JOBCAT 'EMPLOYMENT CATEGORY' /
MINORITY 'MINORITY CLASSIFICATION'.
VALUE LABELS SEX 0 'MALES' 1 'FEMALES' /
JOBCAT 1 'CLERICAL' 2 'OFFICE TRAINEE'
3 'SECURITY OFFICER'
4 'COLLEGE TRAINEE' 5 'EXEMPT EMPLOYEE'
6 'MBA TRAINEE' 7 'TECHNICAL' /
MINORITY 0 'WHITE' 1 'NONWHITE'
MISSING VALUES SALBEG, TIME TO EDLEVEL, JOBCAT (0) / SEX, MINORITY (9).
FORMATS SALBEG SALNOW (COMMA6.0).
COMPUTE LOGBEG=LG10(SALBEG)
COMPUTE LOGNOW=LG10(SALNOW)
REGRESSION VARIABLES=LOGBEG, EDLEVEL, SEX, WORK, MINORITY, AGE /
STATISTICS=R, COEFF, OUTS, F
DEPENDENT=LOGBEG /
METHOD=STEPWISE.
FINISH.
```

- The **DATA LIST** command **defines** the **data file** and variable names and gives the **column locations** for **each** variable.
- The **VARIABLE LABELS** and **VALUE LABELS** commands supply descriptive **labels** for the variables and their **values**.
- The **MISSING VALUES** command assigns the value 0 as missing for **SALBEG**, **TIME**, **TO EDLEVEL**, and **JOBCAT**, and the value 9 as missing for **SEX** and **MINORITY**.
- The **FORMATS** command assigns a **comma** display and-kite format to the variables **SALBEG** and **SALNOW**.

- The COMPUTE commands create the **new variables LOGBEG** and **LOGNOW**.
- The REGRESSION command asks for a **stepwise regression of** the named **variables** with **LOGBEG** as the dependent variable. It **also** asks for the statistics *R*, **COEFF**, **OUTS**, and **F**.

The output from this example is shown in Figure 6.48.