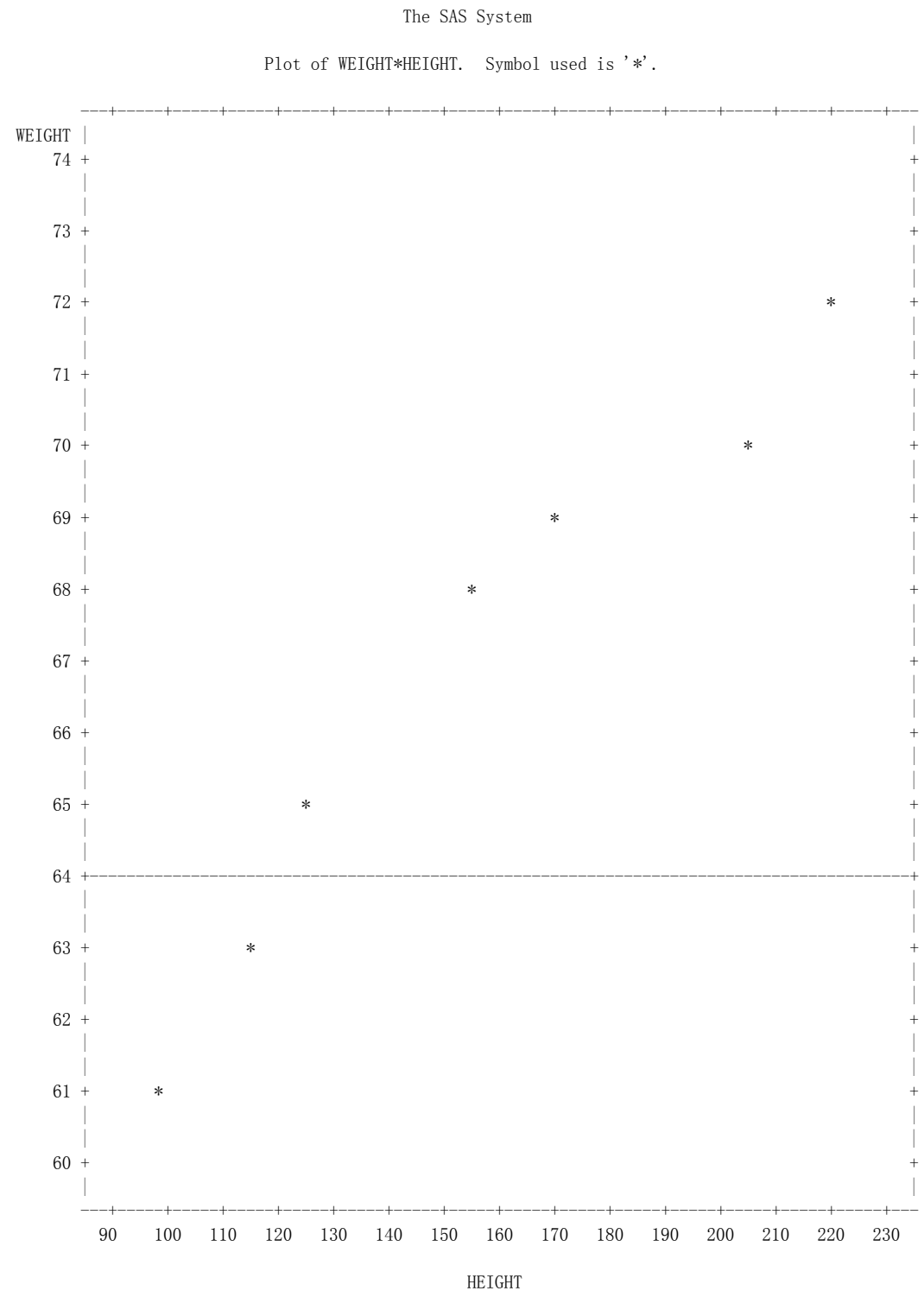
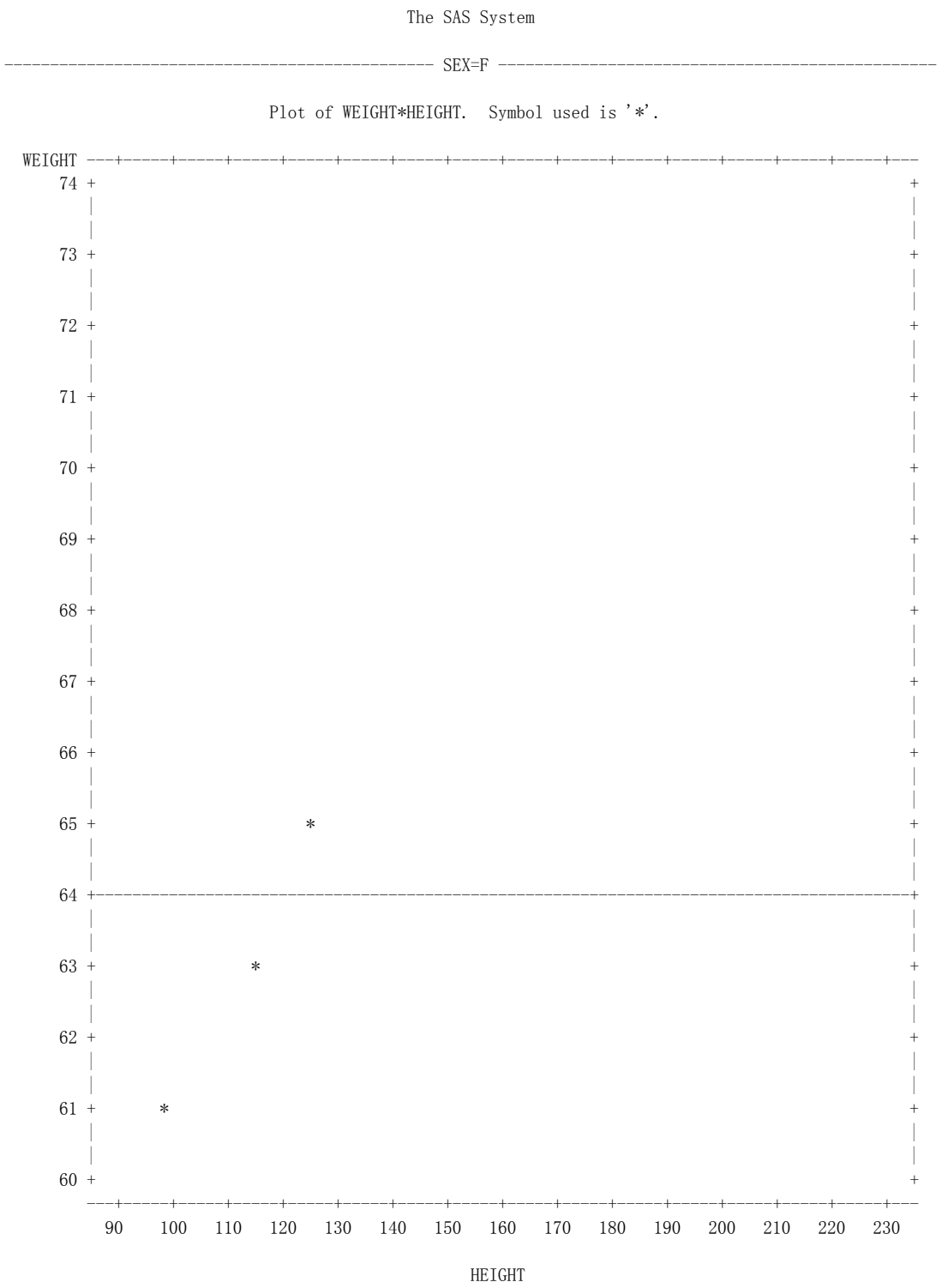
APPENDIX

Topic 5.1: PROC PLOT

Output: Example 5.1(a)

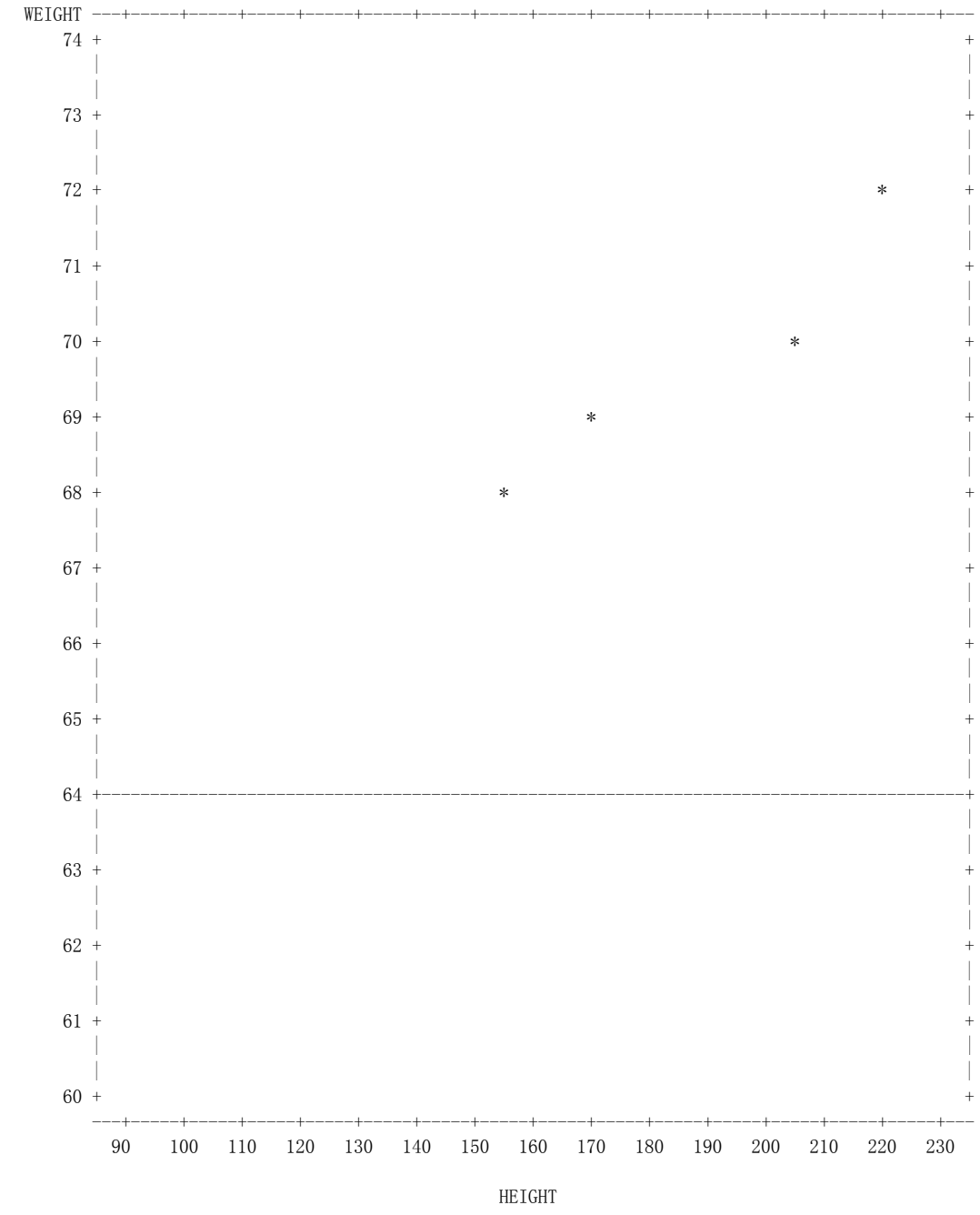


Output: Example 5.1(b)



SEX=M

Plot of WEIGHT*HEIGHT. Symbol used is '*'.



Topic 5.3: PROC MEANS

A list of the commonly requested options for PROC MEANS is shown as below.

Option	**Description**
N	Number of observations on which the statistic was computed
NMISS	Number of missing observations
MEAN	Arithmetic mean
STD	Standard deviation
STDERR	Standard error
MIN	Minimum
MAX	Maximum
SUM	Sum
VAR	Variance
CV	Coefficient of variation
SKEWNESS	Skewness
KURTOSIS	Kurtosis
T	Student's *t*-test whether the population mean is zero
PRT	The probability of obtaining a larger absolute value of *t* (*p*-value)
MAXDEC = n	Where n specifies the number of decimal places for printed statistics
CLM	Two-sided confidence limit
ALPHA	Value of α

Topic 5.4: PROC UNIVARIATE

Some PROC statement options:

Option	Description
ALPHA =	option specifies anα for calculating $(1-\alpha)$ 100% confidence intervals, the default being .05
NORMAL	option requests tests for normality. computed test statistics and *p*-values for the Shapiro-Wilk test (for sample sizes less than or equal to 2000), the Kolmogorov-Smirnov test, the Anderson-Darling test, and the Cram*e*r-von Mises test are output
PLOT	requests that low-resolution descriptive graphics (a histogram or a stem-and-leaf plot, a box plot, and a normal probability plot) are to be produced.

Topic 5.5: PROC FREQ

- Some TABLES options are shown as below.

Option	Description
NOCOL	suppresses printing the column percentage for each cell
NOCUM	suppresses printing the cumulative frequencies and cumulative percentages in one-way frequency tables and in list format
NOROW	suppresses printing the row percentage for each cell
NOPERCENT	suppresses printing the percentage, row percentage, and column percentage in two-way tables, or percentages and cumulative percentages in one-way tables and in list format
NOPRINT	suppresses printing the frequency table but displays other statistics
LIST	prints multi-way tables in list format
BINOMIAL	requests binomial proportion, confidence limits and test for oneway tables
TESTF=	specifies expected frequencies for a one-way table chi-square test
TESTP=	specifies expected proportions for a one-way table chi-square test
CHISQ	requests chi-square tests and measures of association based on chi-square
CELLCHI2	prints each cell's contribution to the total Pearson chi-square statistic
DEVIATION	prints the deviation of the cell frequency from the expected value for each cell.
EXPECTED	prints the expected cell frequency for each cell under the null
FISHER	requests Fisher's exact test for tables larger than 2 ×2
MEASURES	requests measures of association and their asymptotic standard errors
CL	requests confidence limits for the measures statistics
ALPHA=	sets the confidence level for confidence limits
AGREE	requests tests and measures of classification agreement
MISSING	includes missing value in frequency tables

Topic 5.6: PROC CORR

Example 5.6:
Each student recorded three values: test score, the number of hours spent watching TV in the week prior to the test and the number of hours exercising in the week.

```
DATA works;
INPUT score TV exercise @@;
CARDS;
56 6 2 78 7 4 84 5 5 73 4 0 65 8 2
82 4 4 78 6 4 90 5 4 49 7 2 69 5 4
73 4 4 89 6 4 68 8 3 88 8 5 78 5 6
63 5 4 90 4 6 74 5 4 92 5 7 55 7 4
36 9 2 46 7 2 76 8 3 28 9 1 47 8 4
;
PROC CORR DATA=works;
VAR TV exercise;
WITH score;
RUN;
```

The CORR Procedure

1 With Variables: score
2 Variables: TV exercise

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
score	25	69.08000	17.69633	1727	28.00000	92.00000
TV	25	6.20000	1.63299	155.00000	4.00000	9.00000
exercise	25	3.60000	1.60728	90.00000	0	7.00000

Pearson Correlation Coefficients, N = 25
Prob > |r| under H0: Rho=0

	TV	exercise
score	-0.61048	0.66039 ← r
	0.0012	0.0003 ← p-value

- Both hours of TV and hours of exercise are correlated with the test score but exercise is positively correlated while TV is negatively correlated.

- Students who watched more TV tended to have lower scores while the students who spent more time for exercising tended to have higher scores.

Topic 6.1: PROC TTEST

Example 6.1:

Data

FEMALE 120 122 100 118 225 130 136 145 150 142 130 140 90 230 240
MALE 85 146 150 152 165 166 160 162 164 170 180 88 179 190 220

Program

```
DATA NO1;
INPUT GROUP $ @;
DO I=1 TO 10;
INPUT WEIGHT @;
OUTPUT;
END;
CARDS;
FEMALE  120  122  100  118  225  130  136  145  150  142  130  140  90  230
240
MALE  85  146  150  152  165  166  160  162  164  170  180  88  179  190  220
;
PROC TTEST ALPHA=0.1;
CLASS GROUP;
VAR WEIGHT;
RUN;
```

Output

The TTEST Procedure

Statistics

Variable	GROUP	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev	Std Err
WEIGHT	FEMALE	10	119.22	138.8	158.38	24.63	33.77	55.558	10.679
WEIGHT	MALE	10	137.64	152	166.36	18.073	24.779	40.766	7.8358
WEIGHT	Diff (1-2)		-36.17	-13.2	9.7683	23.387	29.618	41.006	13.245

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
WEIGHT	Pooled	Equal	18	-1.00	0.3322
WEIGHT	Satterthwaite	Unequal	16.5	-1.00	0.3334

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
WEIGHT	Folded F	9	9	1.86	0.3700

Topic 6.2: PROC REG

List of options for MODEL statement:

Option	Description
ADJRSQ	computes adjusted R^2
AIC	computes Akaike's information criterion
B	computes parameter estimates for each model
BIC	computes Sawa's Bayesian information criterion
CP	computes Mallows' C_p statistic
MSE	computes MSE for each model
RMSE	displays root MSE for each model
SSE	computes error sum of squares for each model
I	displays inverse of sums of squares and crossproducts
XPX	displays sums-of-squares and crossproducts matrix
CORRB	displays correlation matrix of estimates. This is the $(\mathbf{X}'\mathbf{X})^{-1}$ matrix scaled to unit diagonals.
COVB	displays covariance matrix of estimates. This matrix is $(\mathbf{X}'\mathbf{X})^{-1}s^2$, where s^2 is the estimated mean squared error.
CLB	computes $100(1-\alpha)$% confidence limits for the parameter estimates
CLI	computes $100(1-\alpha)$% confidence limits for an individual predicted value
CLM	computes $100(1-\alpha)$% confidence limits for the expected value of the dependent variable
ALPHA=	sets significance value for confidence and prediction intervals and tests
NOPRINT	suppresses display of results

Example 6.2:

Consider the following data set:

M 155 68 F 99 61 F 115 63 M 205 70 M 170 69
F 125 65 M 220 72 M 180 75 F 145 68 F 160 70
M 150 60 M 172 70 M 180 80 F 145 70 F 155 62
F 120 58

Write a complete SAS program for the data above using PROC REG in order to perform a simple linear regression.

Program

```
TITLE 'EXAMPLE ON PROC REG';
OPTIONS PS=80 ;
DATA reg;
INPUT gender $ ht wt @@;
CARDS;
M 155 68  F 99 61 F 115 63 M 205 70 M 170 69
F 125 65  M 220 72 M 180 75 F 145 68 F 160 70
M 150 60 M 172 70 M 180 80 F 145 70 F 155 62
F 120 58
;
PROC REG CORR;
MODEL ht = wt;
RUN;
```

Output

EXAMPLE ON PROC REG
The REG Procedure

Correlation

Variable	wt	ht
wt	1.0000	0.6955
ht	0.6955	1.0000

The REG Procedure
Model: MODEL1
Dependent Variable: ht

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	7538.89364	7538.89364	13.12	0.0028
Error	14	8045.10636	574.65045		
Corrected Total	15	15584			

Root MSE	23.97187	R-Square	0.4838
Dependent Mean	156.00000	Adj R-Sq	0.4469
Coeff Var	15.36658		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-104.80210	72.25339	-1.45	0.1690
wt	1	3.86016	1.06575	3.62	0.0028

Example 6.3:

Consider the following data set:

Y	X1	X2	X3
2.2	1	0.75	2.50
3.8	2	1.50	2.35
6.1	3	0.85	2.00
8.2	4	1.85	1.85
9.9	5	2.30	1.35
11.8	6	2.90	1.00
14.4	7	3.50	0.85
16.0	8	4.00	0.65
18.0	9	4.45	0.40
19.5	10	5.20	0.25
22.2	11	5.70	0.20
25.0	12	6.10	0.05

Execute the following program using the procedure below,

```
PROC REG;
MODEL Y = X1 X2 X3 / I  XPX  COVB  CLM CLI;
```

Topic 6.3: PROC ANOVA

MEANS Option	Description
alpha=p	This is the significance level used in multiple comparisons. The default is .05.
bon	This requests Bonferroni t-tests of differences between means.
cldiff	This option requests confidence intervals for all pairwise differences between means.
duncan	This option requests Duncan's multiple comparisons.
lines	This option lists the means in descending order, indicating those means that are not significantly different with stars simulating a line segment beside them.
scheffe	This option requests Scheffé's multiple comparisons.
snk	This option requests the Student-Newman-Keuls multiple range test.
lsd	This option performs pairwise t-tests, which is equivalent to Fisher's least-significant-difference test when cell sizes are equal.
tukey	This option performs Tukey's studentized range test.

Example 6.4:

Program

```
DATA SLR;
INPUT FACA $ @;
DO I=1 TO 8;
INPUT Y @;
OUTPUT;
END;
CARDS;
FERT1  12.8  12.6  12.6  12.4  12.8  12.6  11.8  13.0
FERT2 8.2   8.4   8.8    8.4   8.6   8.8   8.7   8.4
FERT3  11.8  13.0  12.8  12.6  12.4  12.6  12.4  12.6
FERT4  16.6  16.2  16.4  16.6  16.8  16.2  16.6  16.5
;
PROC ANOVA;
CLASS FACA;
MODEL Y = FACA;
MEANS FACA / CLM  CLDIFF  LSD  LINES;
RUN;
```

Output

EXAMPLE ON PROC REG

The ANOVA Procedure

Class Level Information

Class	Levels	Values
FACA	4	FERT1 FERT2 FERT3 FERT4

Number of observations 32

The ANOVA Procedure

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	252.8312500	84.2770833	968.11	<.0001
Error	28	2.4375000	0.0870536		
Corrected Total	31	255.2687500			

R-Square	Coeff Var	Root MSE	Y Mean
0.990451	2.354501	0.295048	12.53125

Source	DF	Anova SS	Mean Square	F Value	Pr > F
FACA	3	252.8312500	84.2770833	968.11	<.0001

The ANOVA Procedure

t Confidence Intervals for Y

Alpha	0.05
Error Degrees of Freedom	28
Error Mean Square	0.087054
Critical Value of t	2.04841
Half Width of Confidence Interval	0.21368

FACA	N	Mean	95% Confidence Limits	
FERT4	8	16.4875	16.2738	16.7012
FERT1	8	12.5750	12.3613	12.7887
FERT3	8	12.5250	12.3113	12.7387
FERT2	8	8.5375	8.3238	8.7512

t Tests (LSD) for Y

NOTE: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	28
Error Mean Square	0.087054
Critical Value of t	2.04841
Least Significant Difference	0.3022

Comparisons significant at the 0.05 level are indicated by ***.

FACA Comparison	Difference Between Means	95% Confidence Limits		
FERT4 - FERT1	3.9125	3.6103	4.2147	***
FERT4 - FERT3	3.9625	3.6603	4.2647	***
FERT4 - FERT2	7.9500	7.6478	8.2522	***
FERT1 - FERT4	-3.9125	-4.2147	-3.6103	***
FERT1 - FERT3	0.0500	-0.2522	0.3522	
FERT1 - FERT2	4.0375	3.7353	4.3397	***
FERT3 - FERT4	-3.9625	-4.2647	-3.6603	***
FERT3 - FERT1	-0.0500	-0.3522	0.2522	
FERT3 - FERT2	3.9875	3.6853	4.2897	***
FERT2 - FERT4	-7.9500	-8.2522	-7.6478	***
FERT2 - FERT1	-4.0375	-4.3397	-3.7353	***
FERT2 - FERT3	-3.9875	-4.2897	-3.6853	***

The ANOVA Procedure

t Tests (LSD) for Y

NOTE: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	28
Error Mean Square	0.087054
Critical Value of t	2.04841
Least Significant Difference	0.3022

Means with the same letter are not significantly different.

t Grouping	Mean	N	FACA
A	16.4875	8	FERT4
B	12.5750	8	FERT1
B			
B	12.5250	8	FERT3
C	8.5375	8	FERT2

Example 6.5:

Data Layout

FACTOR2 / FACTOR1	B1	B2	B3
A1	32, 34, 32, 34	26, 36, 38, 32	30, 30, 38, 34
A2	40, 40, 42, 44	66, 64, 62, 66	42, 46, 40, 38

TWOWAY.DAT

A1 B1 32 34 32 34 B2 26 30 28 24 B3 30 30 38 34
A2 B1 40 40 42 44 B2 66 64 62 66 B3 42 46 40 38

A completely randomized two way full factorial model

$$y_{ijk} = \mu + \tau_i + \beta_j + \tau\beta_{ij} + \varepsilon_{ijk}, \varepsilon_{ijk} \sim N(0, \sigma^2)$$

is assumed for the data in TWOWAY.DAT. Using the data given, run PROC ANOVA and the appropriate options.