

```

1 *****;
2 *** EXST7034 Multiple Regression Example ***;
3 *** Problem from Neter, Kutner, Nachtsheim & Wasserman 1996, #6.18 ***;
4 *****;

```

```

5
6 OPTIONS LS=99 PS=80 NOCENTER NODATE NONUMBER;
7
8 DATA ONE; INFILE CARDS MISSOEVER;
9 TITLE1 'EXST7034 - NKNW 6.18 : Mathematician salaries';
10 * LABEL X1 = 'Index of publication quality';
11 * LABEL X2 = 'Number of years experience';
12 * LABEL X3 = 'Grant support success';
13 * LABEL Y = 'Thousands of dollars';
14 INPUT Y X1 X2 X3;
15 CARDS;

```

NOTE: The data set WORK.ONE has 24 observations and 4 variables.

NOTE: DATA statement used:  
real time 0.14 seconds  
cpu time 0.00 seconds

```

40 ;
49 PROC REG DATA=ONE; TITLE2 'Multiple Regression Example 2';
50 MODEL Y = X2 X2X3SUM; RUN;

```

NOTE: 24 observations read.  
NOTE: 24 observations used in computations.  
NOTE: The PROCEDURE REG printed page 11.  
NOTE: PROCEDURE REG used:  
real time 0.09 seconds  
cpu time 0.07 seconds

EXST7034 - NKNW 6.18 : Mathematician salaries  
Multiple Regression Example 2

The REG Procedure  
Model: MODEL1  
Dependent Variable: Y

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	627.38353	313.69176	106.46	<.0001
Error	21	61.87647	2.94650		
Corrected Total	23	689.26000			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	17.89290	1.95684	9.14	<.0001
X2	1	0.31865	0.03556	8.96	<.0001
X2X3SUM	1	1.20345	0.18912	6.36	<.0001

```

51 PROC REG DATA=ONE; TITLE2 'Multiple Regression Example 3';
52 MODEL Y = X1 X2 X3;
53 restrict x1=x3;
54 RUN;

```

NOTE: 24 observations read.  
NOTE: 24 observations used in computations.  
NOTE: The PROCEDURE REG printed page 12.  
NOTE: PROCEDURE REG used:  
real time 0.07 seconds  
cpu time 0.07 seconds

EXST7034 - NKNW 6.18 : Mathematician salaries  
Multiple Regression Example 3

The REG Procedure

Model: MODEL1

Dependent Variable: Y

NOTE: Restrictions have been applied to parameter estimates.

Analysis of Variance		Sum of	Mean		
Source	DF	Squares	Square	F Value	Pr > F
Model	2	627.38353	313.69176	106.46	<.0001
Error	21	61.87647	2.94650		
Corrected Total	23	689.26000			

Root MSE	1.71654	R-Square	0.9102
Dependent Mean	39.50000	Adj R-Sq	0.9017
Coeff Var	4.34566		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	17.89290	1.95684	9.14	<.0001
X1	1	1.20345	0.18912	6.36	<.0001
X2	1	0.31865	0.03556	8.96	<.0001
X3	1	1.20345	0.18912	6.36	<.0001
RESTRICT	-1	-2.33286	6.08223	-0.38	0.7111*

\* Probability computed using beta distribution.

```
55 PROC REG DATA=ONE; TITLE2 'Multiple Regression Example 3';
56 MODEL Y = X1 X2 X3;
57 test x1=x3;
58 RUN;
```

NOTE: 24 observations read.

NOTE: 24 observations used in computations.

The SAS System

NOTE: The PROCEDURE REG printed pages 13-14.

NOTE: PROCEDURE REG used:

real time	0.08 seconds
cpu time	0.08 seconds

Analysis of Variance

Analysis of Variance		Sum of	Mean		
Source	DF	Squares	Square	F Value	Pr > F
Model	3	627.81700	209.27233	68.12	<.0001
Error	20	61.44300	3.07215		
Corrected Total	23	689.26000			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	17.84693	2.00188	8.92	<.0001
X1	1	1.10313	0.32957	3.35	0.0032
X2	1	0.32152	0.03711	8.66	<.0001
X3	1	1.28894	0.29848	4.32	0.0003

Test 1 Results for Dependent Variable Y

Test 1 Results for Dependent Variable Y		Mean			
Source	DF	Square	F Value	Pr > F	
Numerator	1	0.43347	0.14	0.7111	
Denominator	20	3.07215			

```
PROC REG DATA=ONE; TITLE2 'Multiple Regression Example';
MODEL Y = X1 X2 X3 / partial all; RUN;
```

NOTE: 24 observations read.  
NOTE: 24 observations used in computations.  
NOTE: The PROCEDURE REG printed pages 1-9.  
NOTE: PROCEDURE REG used:  
real time 0.26 seconds  
cpu time 0.13 seconds

Analysis of Variance Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	627.81700	209.27233	68.12	<.0001
Error	20	61.44300	3.07215		
Corrected Total	23	689.26000			

Parameter Estimates								
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t	Type I SS	Type II SS	Standardized Estimate
Intercept	1	17.84693	2.00188	8.92	<.0001	37446	244.17168	0
X1	1	1.10313	0.32957	3.35	0.0032	306.73233	34.41851	0.26023
X2	1	0.32152	0.03711	8.66	<.0001	263.79445	230.62548	0.65915
X3	1	1.28894	0.29848	4.32	0.0003	57.29022	57.29022	0.30694

Parameter Estimates								
Variable	DF	Squared Semi-partial Corr Type I	Squared Partial Corr Type I	Squared Semi-partial Corr Type II	Squared Partial Corr Type II	Tolerance	Variance Inflation	
Intercept	1	.	.	.	.	.	0	
X1	1	0.44502	0.44502	0.04994	0.35904	0.73736	1.35619	
X2	1	0.38272	0.68961	0.33460	0.78963	0.77010	1.29852	
X3	1	0.08312	0.48251	0.08312	0.48251	0.88225	1.13347	

Parameter Estimates			
Variable	DF	95% Confidence Limits	
Intercept	1	13.67109 22.02277	
X1	1	0.41565 1.79061	
X2	1	0.24411 0.39893	
X3	1	0.66632 1.91156	

**Example of regression with uncorrelated Xi variables**

```
DATA ONE; INFILE CARDS MISSOVER;
  TITLE1 'EXST7034 - Example NWK Table 8.7 : Uncorrelated variables';
  LABEL Y = 'Crew Productivity Score';
  INPUT TRIAL CREWSIZE BONUSPAY Y;
CARDS; RUN;
  1  4  2  42
  2  4  2  39
  3  4  3  48
  4  4  3  51
  5  6  2  49
  6  6  2  53
  7  6  3  61
  8  6  3  60
;
PROC REG DATA=ONE; TITLE2 'All models in PROC REG';
  MODEL Y = BONUSPAY;
  MODEL Y = CREWSIZE;
  MODEL Y = CREWSIZE BONUSPAY / SS2; RUN;
```

Model: MODEL1  
Dependent Variable: Y Crew Productivity Score

Analysis of Variance	Sum of	Mean			
Source	DF	Squares	Square	F Value	Prob>F
Model	1	171.12500	171.12500	4.128	0.0885
Error	6	248.75000	41.45833		
C Total	7	419.87500			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob >  T	Variable Label
INTERCEP	1	27.250000	11.60773808	2.348	0.0572	Intercept
BONUSPAY	1	9.250000	4.55292946	2.032	0.0885	

Model: MODEL2  
Dependent Variable: Y Crew Productivity Score

Analysis of Variance	Sum of	Mean			
Source	DF	Squares	Square	F Value	Prob>F
Model	1	231.12500	231.12500	7.347	0.0351
Error	6	188.75000	31.45833		
C Total	7	419.87500			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob >  T	Variable Label
INTERCEP	1	23.500000	10.11135912	2.324	0.0591	Intercept
CREWSIZE	1	5.375000	1.98300067	2.711	0.0351	

Model: MODEL3  
Dependent Variable: Y Crew Productivity Score

Analysis of Variance	Sum of	Mean			
Source	DF	Squares	Square	F Value	Prob>F
Model	2	402.25000	201.12500	57.057	0.0004
Error	5	17.62500	3.52500		
C Total	7	419.87500			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob >  T	Type II SS	Variable Label
INTERCEP	1	0.375000	4.74045093	0.079	0.9400	0.022059	Intercept
CREWSIZE	1	5.375000	0.66379590	8.097	0.0005	231.125000	
BONUSPAY	1	9.250000	1.32759180	6.968	0.0009	171.125000	

### Examples of Multicollinearity

```
DATA TWO; INFILE CARDS MISSOVER;
  TITLE1 'EXST7034 - Example NWK Table 8.8 : Perfectly correlated
variables';
  INPUT CASE X1 X2 Y;
CARDS; RUN;
1 2 6 23
2 8 9 83
3 6 8 63
4 10 10 103
;
```

```
PROC REG DATA=TWO; TITLE2 'Generic example';
  MODEL Y = X1;
  MODEL Y = X2;
  MODEL Y = X1 X2 / SS2; RUN;
```

```
DATA TWO; INFILE CARDS MISSOVER;
  TITLE1 'EXST7034 - Modified example NWK Table 8.8 : Perfectly correlated
independent variables';
  INPUT CASE X1 X2 Y;
CARDS; RUN;
1 2 6 23
2 8 12 83
3 7 11 63
4 10 14 103
;
```

```
PROC REG DATA=TWO; TITLE2 'Modified generic example';
  MODEL Y = X1;
  MODEL Y = X2;
  MODEL Y = X1 X2 / SS2; RUN;
```

Generic example

Model: MODEL1                      Dependent Variable: Y  
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	3500.00000	3500.00000	.	.
Error	2	0.00000	0.00000		
C Total	3	3500.00000			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob >  T
INTERCEP	1	3.000000	0.00000000	.	.
X1	1	10.000000	0.00000000	.	.

Generic example

Model: MODEL2                      Dependent Variable: Y  
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	3500.00000	3500.00000	.	.
Error	2	0.00000	0.00000		
C Total	3	3500.00000			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob >  T
INTERCEP	1	-97.000000	0.00000000	.	.
X2	1	20.000000	0.00000000	.	.

Generic example

Model: MODEL3                      Dependent Variable: Y  
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	3500.00000	3500.00000	.	.
Error	2	0.00000	0.00000		
C Total	3	3500.00000			

NOTE: Model is not full rank. Least-squares solutions for the parameters are not unique. Some statistics will be misleading. A reported DF of 0 or B means that the estimate is biased. The following parameters have been set to 0, since the variables are a linear combination of other variables as shown.

$$X2 = +5.0000 * INTERCEP + 0.5000 * X1$$

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob >  T	Type II SS
INTERCEP	B	3.000000	0.00000000	.	.	6.176471
X1	B	10.000000	0.00000000	.	.	3500.000000
X2	0	0	0.00000000	.	.	.

EXST7034 - Modified NWK Table 8.8 : Perfectly correlated independent var  
Modified generic example

Model: MODEL1                      Dependent Variable: Y  
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	3425.17986	3425.17986	91.558	0.0107
Error	2	74.82014	37.41007		
C Total	3	3500.00000			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob >  T
INTERCEP	1	0.985612	7.64217078	0.129	0.9092
X1	1	9.928058	1.03756871	9.569	0.0107

Modified generic example

Model: MODEL2                      Dependent Variable: Y  
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	3425.17986	3425.17986	91.558	0.0107
Error	2	74.82014	37.41007		
C Total	3	3500.00000			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob >  T
INTERCEP	1	-38.726619	11.56551739	-3.348	0.0788
X2	1	9.928058	1.03756871	9.569	0.0107

Modified generic example

Model: MODEL3                      Dependent Variable: Y  
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	3425.17986	3425.17986	91.558	0.0107
Error	2	74.82014	37.41007		
C Total	3	3500.00000			

NOTE: Model is not full rank. Least-squares solutions for the parameters are not unique. Some statistics will be misleading. A reported DF of 0 or B means that the estimate is biased.

The following parameters have been set to 0, since the variables are a linear combination of other variables as shown.

$$X2 = +4.0000 * INTERCEP + 1.0000 * X1$$

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob >  T	Type II SS
INTERCEP	B	0.985612	7.64217078	0.129	0.9092	0.622252
X1	B	9.928058	1.03756871	9.569	0.0107	3425.179856
X2	0	0	0.00000000	.	.	.

EXST7034 - Example NWK (1989) Table 8.1  
Multiple regression : Body Fat

Dependent Variable: Y                      Body Fat

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	396.98461	132.32820	21.516	0.0001
Error	16	98.40489	6.15031		
C Total	19	495.38950			

Root MSE	2.47998	R-square	0.8014
Dep Mean	20.19500	Adj R-sq	0.7641
C.V.	12.28017		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob >  T	Type I SS
INTERCEP	1	117.084695	99.78240295	1.173	0.2578	8156.760500
X1	1	4.334092	3.01551136	1.437	0.1699	352.269797
X2	1	-2.856848	2.58201527	-1.106	0.2849	33.168913
X3	1	-2.186060	1.59549900	-1.370	0.1896	11.545902

Variable	DF	Type II SS	Standardized Estimate	Squared Semi-partial Corr Type I	Squared Partial Corr Type I	Squared Semi-partial Corr Type II
INTERCEP	1	8.468159	0.00000000	.	.	.
X1	1	12.704893	4.26370457	0.71109662	0.71109662	0.02564627
X2	1	7.529278	-2.92870065	0.06695522	0.23175644	0.01519870
X3	1	11.545902	-1.56141679	0.02330672	0.10500972	0.02330672

Variable	DF	Squared Partial Corr Type II	Tolerance	Variance Inflation
INTERCEP	1	.	.	0.00000000
X1	1	0.11434540	0.00141075	708.84291418
X2	1	0.07107507	0.00177197	564.34338573
X3	1	0.10500972	0.00955968	104.60600501

Sequential Parameter Estimates

	INTERCEP	X1	X2	X3
	20.195	0	0	0
	-1.496104641	0.8571865102	0	0
	-19.17424564	0.2223525911	0.6594217965	0
	117.08469478	4.3340920083	-2.856847936	-2.186060252

Collinearity Diagnostics

Number	Eigenvalue	Condition Index	Var Prop INTERCEP	Var Prop X1	Var Prop X2	Var Prop X3
1	3.96796	1.00000	0.0000	0.0000	0.0000	0.0000
2	0.02052	13.90482	0.0004	0.0013	0.0000	0.0014
3	0.01151	18.56570	0.0006	0.0002	0.0003	0.0069
4	8.64793E-6	677.37207	0.9990	0.9985	0.9996	0.9917



Values of regression coefficients and standard errors for each of the three independent variables in the Body Fat Example when other variables are in the model (Neter, Wasserman & Kuttner, 1989, Table 8.1).

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error
X1	1	0.857187	0.12878079
X1   X2	1	0.222353	0.30343892
X1   X3	1	1.000585	0.12823209
X1   X2, X3	1	4.334092	3.01551136
X2	1	0.856547	0.11001562
X2   X1	1	0.659422	0.29118728
X2   X3	1	0.850882	0.11244824
X2   X1, X3	1	-2.856848	2.58201527
X3	1	0.199429	0.32662975
X3   X1	1	-0.431442	0.17661556
X3   X2	1	0.096029	0.16139267
X3   X1, X2	1	-2.186060	1.59549900

Correlations among independent variables in Body Fat Example (NWK, 1990).

	Pearson Correlation Coefficients / Prob> R  under Ho: Rho=0/N = 20		
	X1	X2	X3
X1 Triceps skinfold thickness	1.00000	0.92384	0.45778
		0.0001	0.0424
X2 Thigh circumference	0.92384	1.00000	0.08467
	0.0001		0.7227
X3 Midarm circumference	0.45778	0.08467	1.00000
	0.0424	0.7227	

Some regression results and correlations between the dependent variable and the various independent variables in the Body Fat Example (NWK, 1990).

Dependent Variable: X1 Triceps skinfold thickness Root MSE 0.19946	R = 0.9986	r = 0.9993
Dependent Variable: X2 Thigh circumference Root MSE 0.23295	R = 0.9982	r = 0.9991
Dependent Variable: X3 Midarm circumference Root MSE 0.37699	R = 0.9904	r = 0.9952