

Regressions with Correlation Matrices¹

In rare situations (for example if extremely confidential data are involved) data will be provided in the form of a correlation matrix. Following is an example of how to create and use a correlation data matrix. The regression done with the correlation matrix replicates the results from individual-level data.

SAS Program to Create Correlation Matrix Data and Run Regression with it

```
* corr.sas ;
title1 'corr.sas, Show creating and using correlation matrix' ;

libname x 'j:\migdest\saslib' ;

data one; set x.md1970_a ( keep = sex age higrader ) ;

if age ge 25 and age le 65 ;
if sex = 1 then male = 1 ; else male = 0 ;

title2 'Proc Reg on Traditional Data' ;
proc reg ;
model higrader = age male ; run ;

title2 'Proc Corr, and create data set' ;
proc corr outp= x.pcorr1 ;
var higrader age male ; run ;

title2 'Proc Print data set created by Proc Corr' ;
proc print data = x.pcorr1 ; run ;

title2 'Proc Reg with data set created by Proc Corr' ;
proc reg corr data = x.pcorr1 ;
model higrader = age male ; run ;

title2 'Create SAS Export File to use in SPSS' ;
libname sas2spss xport 'j:\migdest\saslib\pcorr1.xpt' ;
proc copy in = x out = sas2spss ; select pcorr1 ;
run ;
```

SPSS Program to Use Correlation Data Created by SAS

```
* Matrix.sps, showing how to use correlation data created by sas.
* It is important to check to make sure no problems occurred during translation.

get sas data = 'j:\migdest\saslib\pcorr1.xpt'.
* Must rename two variables for SPSS .
save outfile = 'j:\migdest\saslib\pcorr1.sav'
  rename v1 = rowtype_
         v2 = varname_ .
get file = 'j:\migdest\saslib\pcorr1.sav' .
REGRESSION matrix in (*)
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT higrader
  /METHOD=ENTER age male .
```

¹Prepared by Patty Glynn, University of Washington, May 25, 2002.

The REG Procedure
Model: MODEL1
Dependent Variable: higrader highest grade, recoded

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1341198	670599	72057.6	<.0001
Error	2.45E6	22799067	9.30643		
Corrected Total	2.45E6	24140264			

Root MSE	3.05064	R-Square	0.0556
Dependent Mean	11.39425	Adj R-Sq	0.0556
Coeff Var	26.77354		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	Intercept	1	14.06581	0.00788	1785.79	<.0001
AGE		1	-0.06325	0.00016846	-375.47	<.0001
male		1	0.20224	0.00390	51.85	<.0001

The CORR Procedure

3 Variables: higrader AGE male

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
higrader	2449822	11.39425	3.13909	27913882	0	18.00000
AGE	2449822	43.77879	11.57036	107250241	25.00000	65.00000
male	2449822	0.48251	0.49969	1182072	0	1.00000

Pearson Correlation Coefficients, N = 2449822
 Prob > |r| under H0: Rho=0

	higrader	AGE	male
higrader highest grade, recoded	1.00000	-0.23350 <.0001	0.03476 <.0001
AGE	-0.23350 <.0001	1.00000	-0.01103 <.0001
male	0.03476 <.0001	-0.01103 <.0001	1.00000

Obs	_TYPE_	_NAME_	higrader	AGE	male
1	MEAN		11.39	43.78	0.48
2	STD		3.14	11.57	0.50
3	N		2449822.00	2449822.00	2449822.00
4	CORR	higrader	1.00	-0.23	0.03
5	CORR	AGE	-0.23	1.00	-0.01
6	CORR	male	0.03	-0.01	1.00

The REG Procedure

Correlation

Variable	Label	AGE	male	higrader
AGE		1.0000	-0.0110	-0.2335
male		-0.0110	1.0000	0.0348
higrader	highest grade, recoded	-0.2335	0.0348	1.0000

The REG Procedure

Model: MODEL1

Dependent Variable: higrader highest grade, recoded

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1341198	670599	72057.6	<.0001
Error	2.45E6	22799067	9.30643		
Corrected Total	2.45E6	24140264			

Root MSE	3.05064	R-Square	0.0556
Dependent Mean	11.39425	Adj R-Sq	0.0556
Coeff Var	26.77354		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	Intercept	1	14.06581	0.00788	1785.79	<.0001
AGE		1	-0.06325	0.00016846	-375.47	<.0001
male		1	0.20224	0.00390	51.85	<.0001

* Matrix.sps, showing how to use correlation data created by sas.
 * It is important to check to make sure no problems occurred during translation.

```
get sas data = 'j:\migdest\saslib\pcorr1.xpt'.
* Must rename two variables for SPSS .
save outfile = 'j:\migdest\saslib\pcorr1.sav'
  rename v1 = rowtype_
        v2 = varname_ .
```

```
get file = 'j:\migdest\saslib\pcorr1.sav' .
REGRESSION matrix in (*)
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT higrader
  /METHOD=ENTER age male .
```

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.236 (a)	.056	.056	3.05064

a Predictors: (Constant), MALE, AGE

ANOVA (b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1341197.883	2	670598.942	72057.600	.000 (a)
	Residual	22799066.613	2449819	9.306		
	Total	24140264.496	2449821			

a Predictors: (Constant), MALE, AGE

b Dependent Variable: HIGRADER highest grade, recoded

Coefficients (a)

		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
Model		B	Std. Error	Beta		
1	(Constant)	14.066	.008		1785.791	.000
	AGE	-6.325E-02	.000	-.233	-375.473	.000
	MALE	.202	.004	.032	51.847	.000

a Dependent Variable: HIGRADER highest grade, recoded