

1. We wish to explore the relationship between monthly food consumption (y) and family monthly income (x), both measured in hundreds of dollars. Here is part of the output produced by an SPSS analysis of these data:

Regression

Descriptive Statistics

	Mean	Std. Deviation	N
FOOD	10.000000	3.0151134	100
INCOME	25.000000	6.0302269	100

Correlations

		FOOD	INCOME
Pearson Correlation	FOOD	1.000	.667
	INCOME	.667	1.000
Covariance	FOOD	9.091	12.121
	INCOME	12.121	36.364
Sig. (1-tailed)	FOOD	.	.000
	INCOME	.000	.
Sum of Squares and Cross-products	FOOD	900.000	1200.000
	INCOME	1200.000	3600.000
N	FOOD	100	100
	INCOME	100	100

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	INCOME ^b	.	Enter

a. All requested variables entered.

b. Dependent Variable: FOOD

- Compute a and b (i.e. $\hat{\alpha}$ and $\hat{\beta}$)
- Estimate how much a family with a monthly income of \$1500 will spend on food.
- Compute the 95% c.i. for b
- Construct the Anova table
- Using the covariances, verify that the reported value for r_{xy} is correct.
- Using the ANOVA table, compute Multiple R^2 .
- Using $\alpha = .01$, test

$$H_0: \beta = 0$$

$$H_A: \beta > 0$$

Also, briefly explain why a 1-tailed alternative probably makes the most sense.

2. Suppose you ask three people the number of airplane trips over 500 miles (y) and the number of automobile trips over 500 miles (x) they took last year, with the following results (the last 3 columns are computed from the first two):

X	Y	XY	X ²	Y ²
0	2	0	0	4
1	4	4	1	16
2	3	6	4	9
T _x = 3	T _y = 9	T _{xy} = 10	T _{x²} = 5	T _{y²} = 29

- Find the least-squares regression line
- Find SSE, SSR, and SST
- What percent in the variability in y is explained by the values of x (i.e. what does r² equal)?
- What is the value of s_{xy}?
- Do a two-tailed test of whether β = 0 (either an F test or a T-test is ok).

3. The following results were obtained from various bivariate regression models. Construct the complete ANOVA Tables.

- SSR = 10, SST = 60, N = 25.
- SSR = 35, N = 50, F = 7.
- r² = .5, SST = 80, N = 30.
- s_e = 5, N = 100, MSR = 100.

4. A study was done on the relationship between Ratings of Stocks (1 = best, 5 = poorest) and prices bid (as a percent of par value). The following output was obtained from an older, less aesthetically appealing but still functional version of SPSS.

a. Fill in the missing entries (1) through (15): (HINT: Rather than working these in the order given, it may be easier if you work things out in the same order we did in the handouts. Keep in mind that several important pieces of information have already been given to you.)

```

* * * *  M U L T I P L E  R E G R E S S I O N  * * * *

      Mean  Std Dev  Variance  Label
BID      99.177    1.463    2.141
RATING    2.900    1.447    2.095

N of Cases =    20

```

Correlation, Covariance, 1-tailed Sig, Cross-Product:

	BID	RATING
BID	1.000	-.804
	2.141	-1.703
	.	.000
	40.688	-32.357
RATING	-.804	1.000
	-1.703	2.095
	.000	.
	-32.357	39.800

Equation Number 1 Dependent Variable.. BID

R Square (1)
 Standard Error (2)

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	(3)	(4)	(5)
Residual	(6)	(7)	(8)

F = (9) Signif F = (10)

----- Variables in the Equation -----

Variable	B	SE B	Beta	T	Sig T
RATING	(11)	(12)	-.804071	(13)	.0000
(Constant)	(14)	.456926		(15)	.0000

- b. Write the least-squares regression equation.
- c. Do a T-test of $H_0: \beta = -1$ vs. $H_A: \beta <> -1$. Use $\alpha = .05$. Also construct the 95% confidence interval, and note whether or not your results are consistent.
- d. If RATING = 2, what is the predicted value for BID?

5. (The following is adapted from Hays, 4th edition, p. 607). In a study of the origins of gender stereotyping of young girls, a random sample of 35 intact families was taken, in which there was an oldest (or only) girl in the 9th grade. The father answered a questionnaire about his interest in sports and received a score X1. The mother answered a similar questionnaire and received a score X2. The physical education instructor of each girl rated her on general athletic ability, and this was used as variable X3. The dependent variable Y was the girl's own score on a questionnaire on interest in sports. For the moment, we will only work with variables X1 and Y.

The program for this problem is called Hw08.sps. Copy this file and run it. Use the printout to help you answer the following questions:

- a. What are the means and standard deviations of X1 and Y?
- b. What is the correlation between X1 and Y?
- c. What does MSR equal?
- d. What value would we predict for a girl whose father had a score of 30?
- e. Using $\alpha = .05$, test the hypothesis
 - $H_0: \beta = 1$
 - $H_A: \beta <> 1$

- f. Using $\alpha = .05$, test the hypothesis
 $H_0: \beta = 1.5$
 $H_A: \beta \neq 1.5$

Note: For (e) and (f), you only need to give a brief explanation to justify your answer.

COMMENT. Problem # 5 is very easy. It is basically testing whether you have mastered some basic computer skills and whether you can read a computer printout.

6. Hw08.sps originally included an early Christmas present. However, the evil grinch came along and disabled one command and replaced two others with nonsensical comments! If you can fix the program and reinsert the missing commands, you will be able to double-check your answers to problems 1, 2 and 4.

Comment. Problem 6 is also very easy. To double-check problem 1 (which requires somewhat different SPSS cards than we have been using) all you have to do is delete the * before the Regression command (the * turns the command into a comment.) For the others, all you have to do is figure out what the appropriate regression commands are. Just looking at the Regression card that is included for problem 5, or other examples from the course packet or the SPSS manual, should make this straightforward.

Some of the SPSS cards used in this program are new or somewhat different than what you are used to (e.g. the Matrix Data and RECODE commands), so see if you can figure out exactly what they are doing and why they work.

7. Use Stata for the following. Make sure your answers are consistent with your earlier results.
- Using the same data as in problem 1, regress food on income. Note that you'll have to use the `corr2data` command to create a pseudo-replication of the data. See the handout on using Stata for OLS regression for details. Rounding error may cause your results to slightly differ from what you got before.
 - Using the raw data presented in problem 2, regress y on x.
 - Using Stata to run the regressions from problems 4 and 5 is OPTIONAL.