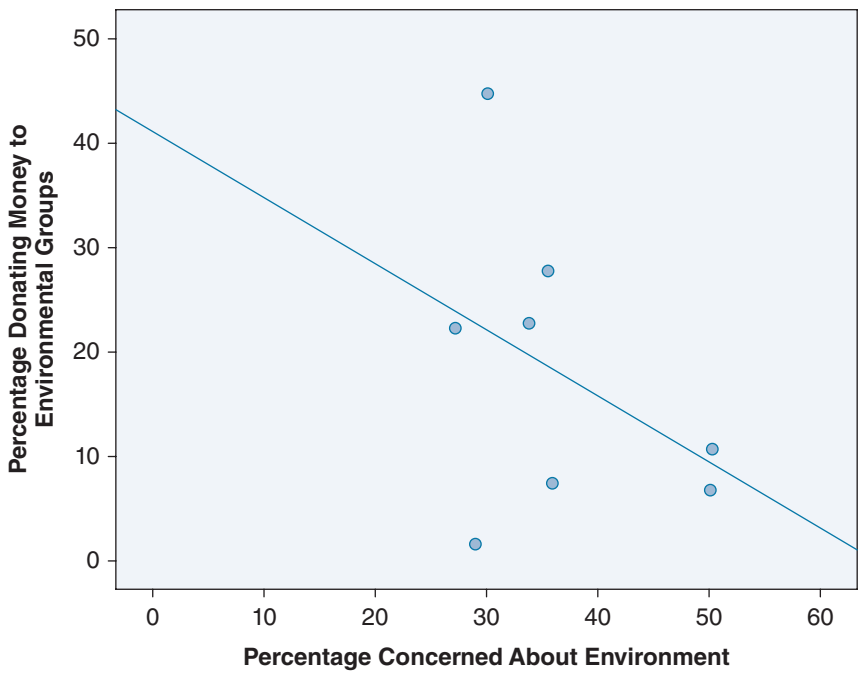
CHAPTER 12— ANSWERS TO EXERCISES

1.

a. On the scatterplot below, the regression line has been plotted to make it easier to see the relationship between the two variables.



b. The scatterplot shows that there is a general linear relationship between the two variables. There is not a lot of scatter about the straight line describing the relationship. As the percentage of respondents concerned about the environment increases, the percentage of respondents donating money to environmental groups decreases.

c. The Pearson correlation coefficient between the two variables is –0.40. This is consistent with the scatterplot that indicated a negative relationship between being concerned about the environment and actually donating money to environmental groups.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Percentage Concerned	Percentage Donating					
Country	X	Y	$(X-\bar{X})$	$(X-\bar{X})^2$	$(Y-\bar{Y})$	$(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$
United States	33.8	22.8	−2.69	7.24	4.77	22.75	−12.83
Austria	35.5	27.8	−0.99	0.98	9.77	95.45	−9.67
The Netherlands	30.1	44.8	−6.39	40.83	26.77	716.63	−171.06
Slovenia	50.3	10.7	13.81	190.72	−7.33	53.73	−101.23
Russia	29.0	1.6	−7.49	56.10	−16.43	269.94	123.06
Philippines	50.1	6.8	13.61	185.23	−11.23	126.11	−152.84
Spain	35.9	7.4	−0.59	0.35	−10.63	113.00	6.27
Denmark	27.2	22.3	−9.29	86.30	4.27	18.23	−39.67
	$\sum X = 291.9$	$\sum Y = 144.2$	−0.02[a]	567.75	0.04[a]	1,415.84	−357.97

$$\text{Mean } X = \bar{X} = \frac{\sum X}{N} = \frac{291.9}{8} = 36.49$$

$$\text{Mean } Y = \bar{Y} = \frac{\sum Y}{N} = \frac{144.2}{8} = 18.03$$

$$\text{Variance } (Y) = s_Y^2 = \frac{\sum (Y-\bar{Y})^2}{N-1} = \frac{1{,}415.84}{7} = 202.26$$

$$\text{Standard deviation } (Y) = s_Y = \sqrt{202.26} = 14.22$$

$$\text{Variance } (X) = s_X^2 = \frac{\sum (X-\bar{X})^2}{N-1} = \frac{567.75}{7} = 81.11$$

$$\text{Standard deviation } (X) = s_X = \sqrt{81.11} = 9.01$$

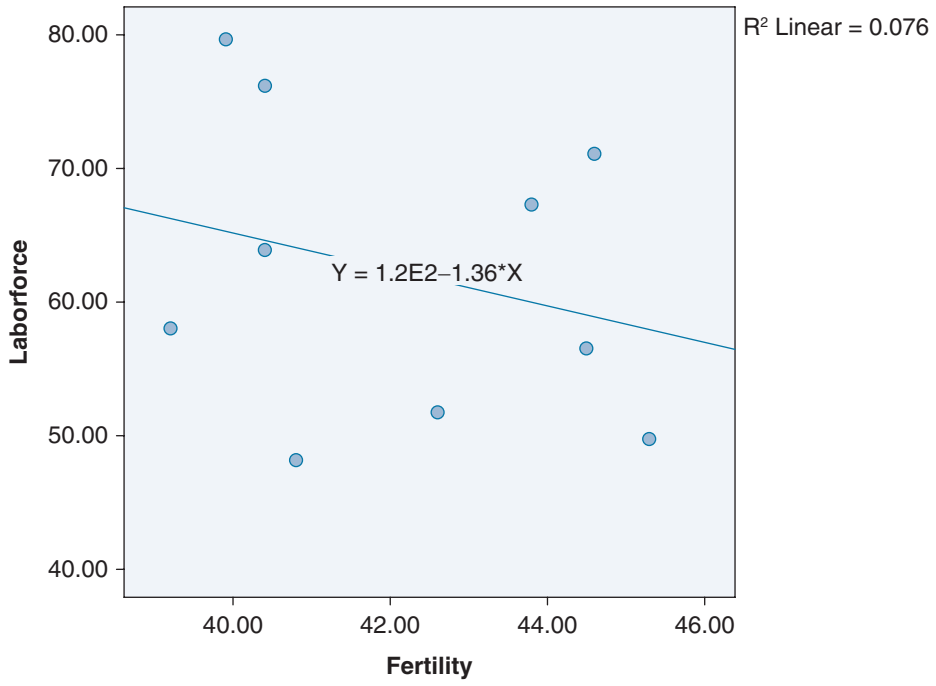
$$\text{Covariance } (X, Y) = s_{XY} = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{N-1} = \frac{-357.97}{7} = -51.14$$

$$r = \frac{s_{XY}}{s_X s_Y} = \frac{-51.14}{(9.01)(14.22)} = -0.40^a$$

Note: Answers may differ slightly due to rounding.

2.

a.



b. The correlation coefficent is –0.275; the coefficient of determination is 0.076.

c. There is a weak negative relationship between the two variables. Adolescent fertility rate can only explain 8% of the variance in predicting labor force participation.

3.

a. The correlation coefficient is –0.45.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Country	GNP per Capita	Percentage Willing to Pay					
State	X	Y	$(X - \bar{X})$	$(X - \bar{X})^2$	$(Y - \bar{Y})$	$(Y - \bar{Y})^2$	$(X - \bar{X})(Y - \bar{Y})$
United States	29.24	44.9	2.72	7.40	−1.64	2.69	−4.46
Ireland	18.71	53.3	−7.81	61.00	6.76	45.70	−52.80
The Netherlands	24.78	61.2	−1.74	3.03	14.66	214.92	−25.51
Norway	34.31	40.7	7.79	60.68	−5.84	34.11	−45.49
Sweden	25.58	32.6	−0.94	0.88	−13.94	194.32	13.10
	$\sum X = 132.62$	$\sum Y = 232.7$	−0.02[a]	132.99	0.04[a]	491.74	−115.16
Mean $X = \bar{X} = \frac{\sum X}{N} = \frac{132.62}{5} = 26.52$							
Mean $Y = \bar{Y} = \frac{\sum Y}{N} = \frac{232.7}{5} = 46.54$							

(Continued)

(Continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Country	GNP per Capita	Percentage Willing to Pay					
State	X	Y	$(X-\bar{X})$	$(X-\bar{X})^2$	$(Y-\bar{Y})$	$(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$
Variance $(X) = s_X^2 = \frac{\sum(X-\bar{X})^2}{N-1} = \frac{132.99}{4} = 33.25$							
Standard deviation $(X) = s_X = \sqrt{33.25} = 5.77$							
Variance $(Y) = s_Y^2 = \frac{\sum(Y-\bar{Y})^2}{N-1} = \frac{491.74}{4} = 122.94$							
Standard deviation $(Y) = S_Y = \sqrt{122.94} = 1.09$							
Covariance $(X, Y) = s_{XY} = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{N-1} = \frac{-115.16}{4} = -28.79$							
$r = \frac{s_{XY}}{s_X s_Y} = \frac{-28.79}{(5.77)(11.09)} = -0.45$[a]							

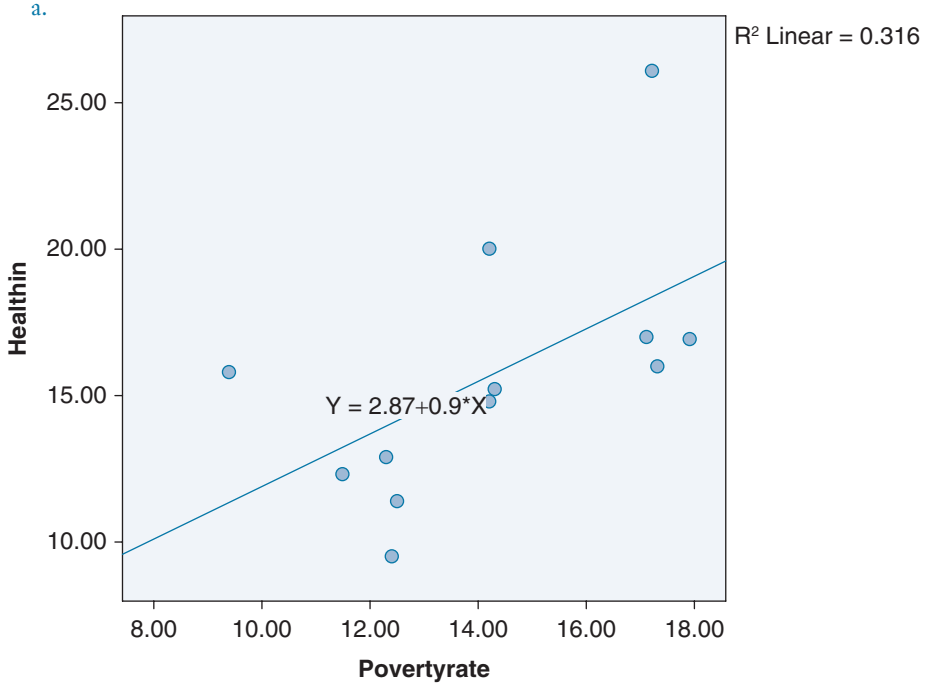
Notes:

a. Answers may differ slightly due to rounding.

b. A correlation coefficient of –0.45 means that relatively high values of GNP are moderately negatively assoicated with low values of percentage of residents willing to pay higher prices to protect the environment.

4.

a.



b. The regression equation is $Y = 2.874 + 0.901(X)$. The data points for Texas, California, New Jersey, and Wisconsin are furthest away from the straight-fitting line.

5. The analysis reveals a negative relationship between years of education and number of children. The bivariate regression equation is $Y = 3.537 + -0.118X$. For each year increase in education, the number of children is predicted to decrease by 0.118. The model explains just 5% of the variance; however, based on the ANOVA F obtained, we can reject the null hypothesis that $r^2 = 0$.

6. The SPSS output confirms the negative relationship between educational attainment and television hours; as years of education increases, television viewing decreases. The F obtained is 31.576 (significant at .000). We can reject the null hypothesis and conclude that the relationship between the two variables is significant.

7.

a. The regression analysis confirms a positive relationship between years of education and total donations given in the past year. The F obtained is 10.578 (significant at .001). We can conclude that the relationship between the two variables is significant.

b. For respondent with 14 years of education: $2043.86

For respondent with 20 years of education: $3868.50

8. The scatterplot reveals a positive relationship between respondent education and mother's education. The bivariate regression equation is $Y = 9.784 + 0.360(X)$. For each year increase in mother's education, respondent's education is predicted to increase by 0.360 years. The model explains 21% of the variance.

9.

a. For males: $Y = 9.768 + 0.355X$

For females: $Y = 9.770 + 0.367X$

b. For males, mother with 20 years of education: $9.768 + 0.355(20) = 16.87$

For females, mother with 20 years of education: $9.770 + 0.367(20) = 17.11$

c. The model for females has a slightly higher r^2. Mother's education explains 22% of the variance in female respondent education compared with the 20% explained for male respondent education. Based on the F-obtained statistic, both models are significant.

10.

a. Both slopes confirm our original hypotheses. Holding number of children constant, for each increase in year of education, number of TV hours is predicted to decrease by 0.164. Holding years of education constant, for each increase in number of children, number of TV hours is predicted to increase by 0.04.

b.

$$\text{Multiple regression equation: } Y = 5.095 + -0.164(X_1) + 0.04(X_2)$$
$$Y = 5.095 + -0.164(16) + 0.04(2) = 2.55 \text{ television hours}$$
$$\text{Bivariate regression equation: } Y = 5.166 + -0.164(X)$$
$$Y = 5.166 + -0.164(16) = 2.54 \text{ television hours}$$

Television hours increases by 0.01 (2.55 – 2.54) with the addition of number of children in the regression model.

c. The amount of explained variance is identical for both models, 4%.

11.

a. Both hypotheses are confirmed.

The slope for education is 0.598. Holding age constant, for each year increase in education, Internet hours per week increases by 0.598.

The slope for age is –0.236. Holding years of education constant, for each year increase in age, Internet hours per week decreases by 0.236.

b. $Y = 14.395 + 0.598(X_1) + -0.236(X_2)$

$Y = 14.395 + 0.598(16) + -0.236(55) = 10.98$ Internet hours per week

c. $Y = 14.395 + 0.263(X_1) + -0.047(X_2)$

Education has the strongest effect on Internet hours per week (beta = .263).

d. The R^2 is 0.065. Education and age explain 6.5% of the variance in predicting Internet hours per week. This is a weak prediction model.

e. The correlation between Internet hours per week and age of respondent is –0.231, indicating a weak negative relationship. The correlation between Internet hours per week and education is 0.088, indicating a weak positive relationship. Finally, the correlation between age and education is –0.009, a weak negative relationship. The only significant correlation is the one between Internet hours and age.

12.

a. Groups 2 and 3 have five significant correlations. First-generation students have two.

b. This correlation is significant for all three groups. It is highest for first-generation students, followed by Group 3 and Group 2. For first-generation students, the correlation of 0.48 indicates a positive moderate relationship between intention and behavior. For Group 3, the correlation of 0.44 indicates a positive moderate relationship between the two variables. For Group 2, the correlation of 0.32 indicates a positive weak relationship between the variables.

c. The correlation for peer support and intention is highest for Group 2 (students with at least one parent with college experience but no degree), which is 0.49.

13.

a. $Y = 3.91 + -0.115(X_1) + -0.038(X_2) + 0.018(X_3) + -0.017(X_4)$

(X_1 = education, X_2 = children, X_3 = age, X_4 = hours worked per week)

Holding all the other independent variables constant,

For each year of increase in education, television viewing should decrease by 0.115 hours.

For each additional child, television viewing decreases by 0.038 hours.

For each additional year of age, television viewing increases by 0.018 hours.

For each additional hour of work, television viewing decreases by 0.017 hours.

b.

Education, –0.202

Hours worked last week, –0.148

Age, 0.139

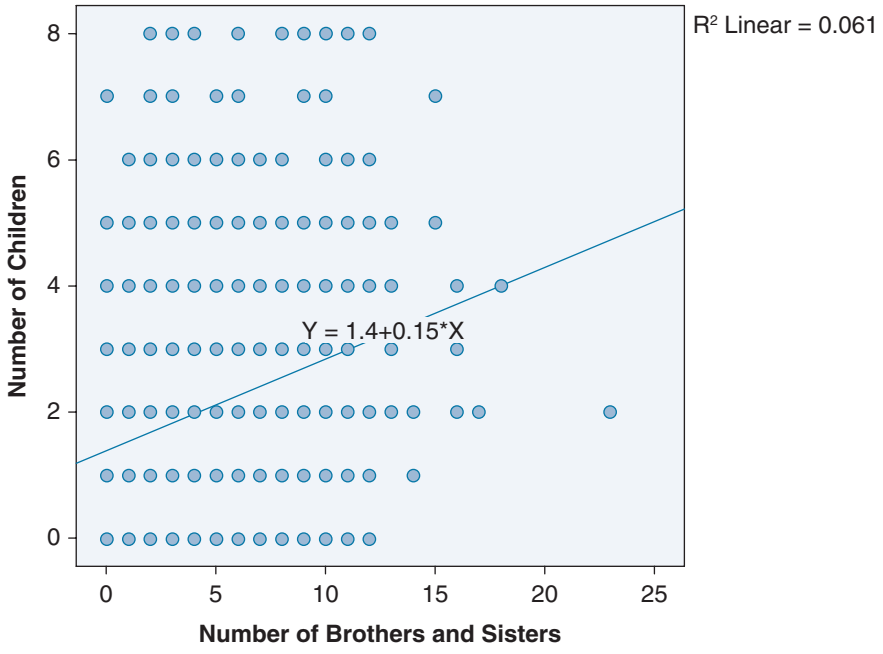
Number of children, –0.034

c. Together these four independent variables reduce the error in predicting TVHOURS by 8.3%. This is a weak prediction model.

SPSS SOLUTIONS

1.

a. As the number of siblings increases, the number of children also increases. There is quite a bit of scatter about the regression line, so there isn't a very strong relationship between the two variables. A linear relationship appears to be a reasonable fit to the relationship. This is a positive relationship.



b. The output from SPSS is omitted. The intercept is 1.396 and the slope is 0.145. The coefficient of determination is 0.061 and the correlation coefficient is 0.247. The positive slope and correlation coefficient confirm the positive relationship between the two variables.

c. Predicted number of children for an individual with 3 siblings:

$$Y = 1.396 + 3(.145) = 1.83$$

d. For someone with 0 siblings, 1.396 (the value of the slope).

2.

a. The slope of the male equation is 0.17 and the slope of the female equation is 0.12. The intercept of the male equation is 1.123 of the female equation, 1.646. The values of the

coefficient of determination for males and females are 0.064 and 0.051, respectively. Note that the intercept for females is higher than for males. The equation for females has a slightly larger coefficient of determination; however, both coefficients of determination suggest weak relationships between the number of siblings a respondent has and his or her number of children.

b. For females:

$$1.646 + 0.12(6) = 2.37$$

For males:

$$1.123 + 0.17(6) = 2.14$$

Based on 6 siblings, the predicted number of children is slightly higher for women.

3.

a. The slope of the white equation and the black equation is 0.142. The intercept of the white equation is 1.426; of the black equation, 1.591. The values of the coefficient of determination for whites and blacks are 0.059 and 0.047, respectively.

The equation for whites has a larger coefficient of determination; however, both coefficients of determination suggest weak relationships between the number of siblings a respondent has and the number of his or her children.

b. For whites:

$$1.426 + 0.142(1) = 1.57$$

$$1.426 + 0.142(4) = 1.99$$

$$1.426 + 0.142(7) = 2.42$$

For blacks:

$$1.591 + 0.142(1) = 1.73$$

$$1.591 + 0.142(4) = 2.16$$

$$1.591 + 0.142(7) = 2.59$$

4.

a.

	Slope	Y Intercept	Coefficient of Determination
Married	0.099	1.756	0.033
Divorced	0.098	1.785	0.046

Note that the slope for married respondents is slightly higher than for divorced respondents, yet the divorced respondents have a higher intercept than the married respondents. The equation for divorced respondents has a larger coefficient of determination; however, both coefficients of determination suggest a weak relationship between the number of siblings a respondent has and the number of his or her children.

b. For married:

$$1.756 + 0.099(1) = 1.86$$

$$1.756 + 0.099(4) = 2.15$$

$$1.756 + 0.099(7) = 2.45$$

For divorced:

$$1.785 + 0.098(1) = 1.88$$

$$1.785 + 0.098(4) = 2.18$$

$$1.785 + 0.098(7) = 2.47$$

c. Overall, on the basis of the coefficients of determination from both analyses, neither equation predicts number of children all that well. The coefficients of determination (r^2) for both married and divorced respondents are very similar and small (0.033 and 0.046).

5.

a. The output from SPSS is omitted. The intercept is 10.333 and the slope is 0.327. The coefficient of determination is 0.205 and the correlation coefficient is 0.452. For every additional year that respondent's father spends in school, we can expect an increase in the respondent's level of education of 0.327 years. When a respondent's father has zero years of education, we can expect the respondent to have 10.33 years of school. When using a respondent's father's education to predict a respondent's education, we improve our prediction by 20%, indicating a moderate positive relationship ($r = 0.452$).

b. The output from SPSS is omitted. The intercept is 9.560 and the slopes are 0.179 and 0.220 for mother's education (MAEDUC) and father's education (PAEDUC), respectively. The coefficient of determination is 0.239 and the correlation coefficient is 0.488. When both a respondent's father and mother have zero years of education, we expect a respondent to have 9.560 years of education, or about 9.6 years of school. Holding a respondent's father's education constant, for every additional year that respondent's mother spends in school, we can expect an increase in the respondent's level of education of 0.220 years, or about 3 months.

Holding a respondent's mother's education constant, for every additional year that respondent's father spends in school, we can expect an increase in the respondent's level of education of 0.179 years, or about 2 months. The value of R^2 is 0.239; thus, 24% of the variation in a respondent's level of education can be explained by taking into account the respondent's father's and mother's level of education.

c. Our prediction improved by taking into account a respondent's mother's level of education. However, that said, our prediction improved only by 3.4% (0.239 – 0.205 = 0.034).

d. Father with 12 years of education:

$$10.333 + 0.327(12) = 14.26$$

Father and mother, each with 12 years of education:

$$9.560 + 0.179\ (12) + 0.220\ (12) = 14.35$$

e. Based on the F ratio of 162.329, we reject the null hypothesis that R^2 is equal to 0. There is a significant relationship between MAEDUC and PAEDUC and EDUC.