Phrasing Results That Are Not Significant

Correlations

			ID	GRADE
Spearman's rho	ID	Correlation Coefficient	1.000	.000
		Sig. (2-tailed)	.	1.000
		N	4	4
	GRADE	Correlation Coefficient	.000	1.000
		Sig. (2-tailed)	1.000	.
		N	4	4

Using our SAMPLE.SAV data set from the previous chapters, we could calculate a Spearman *rho* correlation between ID and GRADE. If so, we would get the output seen to the left. The correlation coefficient equals .000 and has a **significance** level of 1.000. Note that this is rounded up and is not, in fact, 1.000. Thus, we could state the following in a results section:

> A Spearman *rho* correlation coefficient was calculated for the relationship between a subject's ID number and grade. An extremely weak correlation that was not significant was found (r (2) = .000, $p > .05$). ID number is not related to grade in the course.

Practice Exercise

Use Practice Data Set 2 in Appendix B. Determine the strength of the relationship between salary and job classification by calculating the Spearman *rho* correlation.

Section 5.3 Simple Linear Regression

Description

Simple linear regression allows the prediction of one variable from another.

Assumptions

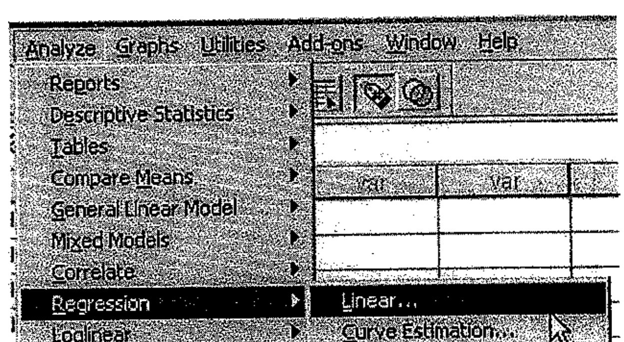
Simple linear regression assumes that both variables are **interval- or ratio-scaled.** In addition, the **dependent variable** should be normally distributed around the prediction line. This, of course, assumes that the variables are related to each other linearly. Normally, both variables should be normally distributed. Dichotomous variables (variables with only two levels) are also acceptable as **independent variables.**

SPSS Data Format

Two variables are required in the SPSS data file. Each subject must contribute to both values.

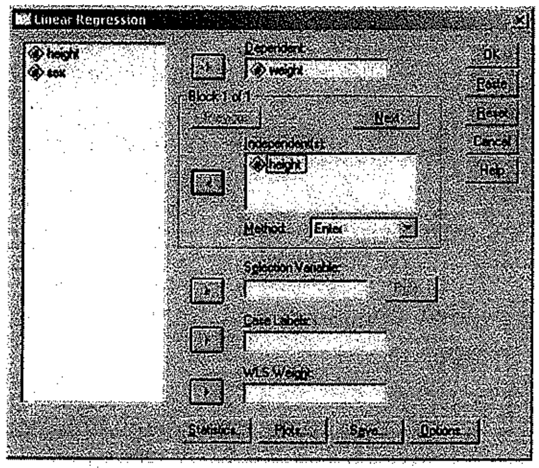
Running the Command

Click <u>A</u>nalyze, then <u>R</u>egression, then <u>L</u>inear. This will bring up the main dialog box for linear regression. On the



left side of the dialog box is a list of the variables in your data file (we are using the HEIGHT.SAV data file from the start of this section). On the right are blocks for the **dependent variable** (the variable you are trying to predict), and the **independent variable** (the variable from which we are predicting).

We are interested in predicting someone's weight based on his or her height. Thus, we should place the variable WEIGHT in the **dependent variable** block and the variable HEIGHT in the **independent variable** block. Then we can click *OK* to run the analysis.



Reading the Output

For simple linear regressions, we are interested in three components of the output. The first is called the Model Summary, and it occurs after the Variables Entered/Removed section. For our example, you should see this output. *R* Square (called the **coefficient of determination**) gives you the proportion of the variance of your **dependent variable** (WEIGHT) that can be explained by variation in your **independent variable** (HEIGHT). Thus, 64.9% of the variation in weight can be explained by differences in height (taller people weigh more).

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.806[a]	.649	.624	16.1480

a. Predictors: (Constant), HEIGHT

The **standard error of estimate** gives you a measure of dispersion for your prediction equation. Using the prediction equation, 68% of the data will fall within one **standard error of estimate** of the predicted value. Just over 95% will fall within two standard errors. Thus, in the example above, 95% of the time, our estimated weight will be within 32.296 pounds of being correct (i.e., $2 \times 16.148 = 32.296$).

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6760.323	1	6760.323	25.926	.000[a]
	Residual	3650.614	14	260.758		
	Total	10410.938	15			

a. Predictors: (Constant), HEIGHT

b. Dependent Variable: WEIGHT

The second part of the output that we are interested in is the ANOVA summary table. More information on reading ANOVA summary tables will be given in Chapter 6. For now, the important number here is the significance level on the far right. If that value is less than .05, then we have a significant linear regression. If it is larger than .05, we do not.

The final section of the output is the table of coefficients. This is where the actual prediction equation can be found.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-234.681	71.552		-3.280	.005
	HEIGHT	5.434	1.067	.806	5.092	.000

a. Dependent Variable: WEIGHT

In most texts, you learn that $Y' = a + bX$ is the regression equation. Y' is your **dependent variable**, and X is your **independent variable**. In SPSS output, the values of both a and b are found in the B column. The first value, –234.681, is the value of a (and labeled Constant). The second value, 5.434, is the value of b (and labeled with the name of the **independent variable**). Thus, our prediction equation for the example above is WEIGHT' = –234.681 + 5.434 (HEIGHT). In other words, the average subject who is an inch taller than another subject weighs 5.434 more pounds. A person who is 60 inches tall should weigh –234.681 + 5.434(60) = 91.359 pounds. Given our earlier discussion of standard error of the estimate, 95% of people who are 60 inches tall will weigh between 59.063 (91.359 – 32.296 = 59.063) and 123.655 (91.359 + 32.296 = 123.655) pounds.

Drawing Conclusions

Conclusions from regression analyses indicate (a) whether or not a significant prediction equation was obtained, (b) the direction of the relationship, and (c) the equation itself.

Phrasing Results That Are Significant

In the example beginning on page 46, we obtained an R Square of .649 and a regression equation of WEIGHT' = –234.681 + 5.434 (HEIGHT). The ANOVA resulted in F = 25.926 with 1 and 14 degrees of freedom. The F is significant at the less than .001 level. Thus, we could state the following in a results section:

> A simple linear regression was calculated predicting subjects' weight based on their height. A significant regression equation was found ($F(1,14)$ = 25.926, $p < .001$), with an R^2 of .649. Subjects' predicted weight is equal to –234.68 + 5.43 (height) pounds when height is measured in inches. Subjects' average weight increased 5.43 pounds for each inch of height.

The conclusion states the direction (increase), strength (.649), value (25.926), degrees of freedom (1,14), and significance level (< .001) of the regression. In addition, a statement of the equation itself is included.

Phrasing Results That Are Not Significant

If the ANOVA is not significant, the section of the output labeled *Sig.* will be greater than .05, and the regression equation is not significant. A results section might include the following statement:

> A simple linear regression was calculated predicting subjects' *ACT* scores based on their height. The regression equation was not significant ($F(1,14) = 1.21$, $p > .05$) with an R^2 of .062. Height cannot be used to predict *ACT* scores.

Note that for results that are not significant, the ANOVA results and R^2 results are given, but the regression equation is not.

Practice Exercise

Use Practice Data Set 2 in Appendix B. If we want to predict salary from years of education, what salary would you predict for someone with 12 years of education? What salary would you predict for someone with a college education (16 years)?

Section 5.4 Multiple Linear Regression

Description

The multiple linear regression analysis allows the prediction of one variable from several other variables.

Assumptions

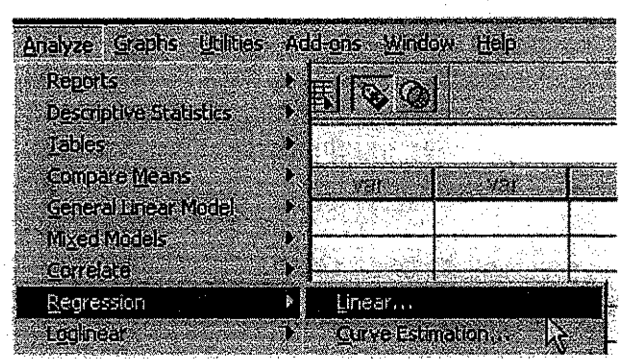
Multiple linear regression assumes that all variables are **interval-** or **ratio-scaled.** In addition, the **dependent variable** should be normally distributed around the prediction line. This, of course, assumes that the variables are related to each other linearly. All variables should be normally distributed. Dichotomous variables are also acceptable as **independent variables.**

SPSS Data Format

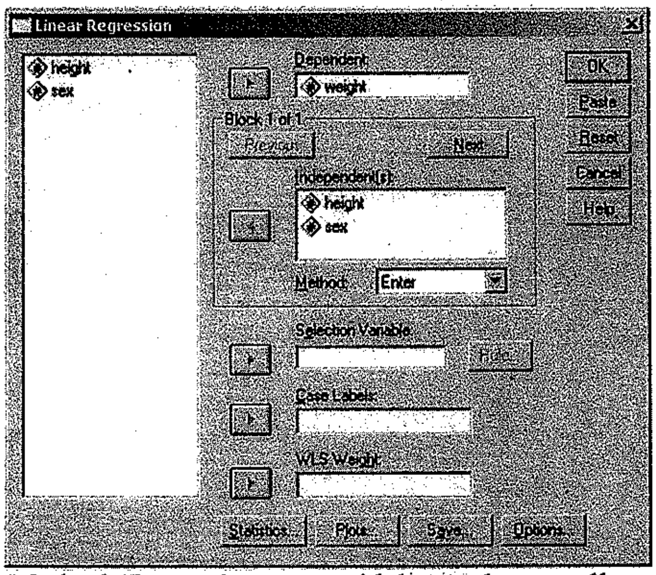
At least three variables are required in the SPSS data file. Each subject must contribute to all values.

Running the Command

Click *Analyze*, then *Regression*, then *Linear*. This will bring up the main dialog box for linear regression. On the left side of the dialog box is a list of the variables in your data file (we are using the HEIGHT.SAV data file from the start of this chapter). On the right side of the dialog box are blanks for the **dependent variable** (the variable you are trying to



predict) and the **independent variables** (the variables from which you are predicting).



We are interested in predicting someone's weight based on his or her height and sex. We believe that both sex and height influence weight. Thus, we should place the **dependent variable** WEIGHT in the *Dependent* block and the **independent variables** HEIGHT and SEX in the *Independent(s)* block. Enter them both in Block 1.

This will perform an analysis to determine if WEIGHT can be predicted from SEX and/or HEIGHT. There are several methods SPSS can use to conduct this analysis. These can be selected with the *Method* box. Method *Enter*, the most widely used, puts all variables in the equation, whether they are significant or not. The other methods use various means to enter only those variables that are significant predictors. Click *OK* to run the analysis.

Reading the Output

For multiple linear regression, there are three components of the output in which we are interested. The first is called the Model Summary, which is found after the Variables Entered/Removed section. For our example, you should get the output to the right. *R* Square (called the **coefficient of determination**) tells you the proportion of the **variance** in the **dependent variable** (WEIGHT) that can be explained by variation in the **independent variables** (HEIGHT and SEX, in this case). Thus, 99.3% of the variation in weight can be explained by differences in height and sex (taller people weigh more, and men weigh more). Note that by adding a second variable, our *R* Square goes up from .649 to .993. The .649 was obtained using the Simple Linear Regression example in Section 5.3.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.997[a]	.993	.992	2.2957

a. Predictors: (Constant), HEIGHT, SEX

The Standard Error of the Estimate gives you a margin of error for the prediction equation. Using the prediction equation, 68% of the data will fall within one standard error of the estimate of the predicted value. Just over 95% will fall within two standard errors of the estimates. Thus, in the example above, 95% of the time, our estimated weight will be within 4.591 (2.296 times 2) pounds of being correct. In our Simple Linear Regression example in Section 5.3, this number was 32.296. Note the higher degree of accuracy.

The second part of the output that we are interested in is the ANOVA summary table. For more information on reading ANOVA tables, refer to the sections on ANOVA in

Chapter 6. For now, the important number is the **significance** level on the far right. If that value is less than .05, then we have a significant linear regression. If it is larger than .05, we do not.

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	10342.424	2	5171.212	981.202	.000[a]
	Residual	68.514	13	5.270		
	Total	10410.938	15			

a. Predictors: (Constant), HEIGHT, SEX

b. Dependent Variable: WEIGHT

The final section of output we are interested in is the table of coefficients. This is where the actual prediction equation can be found.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	47.138	14.843		3.176	.007
	SEX	-39.133	1.501	-.767	-26.071	.000
	HEIGHT	2.101	.198	.312	10.588	.000

a. Dependent Variable: WEIGHT

In most texts, you learn that $Y' = a + bX$ is the regression equation. For multiple regression, our equation changes to $Y' = B_0 + B_1X_1 + B_2X_2 + \dots + B_zX_z$ (where z is the number of Independent Variables). Y' is your **dependent variable**, and the Xs are your **independent variables**. The Bs are listed in a column. Thus, our prediction equation for the example above is WEIGHT' = 47.138 – 39.133(SEX) + 2.101(HEIGHT) (where SEX is coded as 1 = Male, 2 = Female, and HEIGHT is in inches). In other words, the average difference in weight for subjects who are one inch different in height is 2.101 pounds. Males tend to weigh 39.133 pounds more than females. A female who is 60 inches tall should weigh 47.138 – 39.133(2) + 2.101(60) = 94.932 pounds. Given our earlier discussion of the **standard error of estimate**, 95% of females who are 60 inches tall will weigh between 90.341 (94.932 – 4.591 = 90.341) and 99.523 (94.932 + 4.591 = 99.523) pounds.

Drawing Conclusions

Conclusions from regression analyses indicate (a) whether or not a significant prediction equation was obtained, (b) the direction of the relationship, and (c) the equation itself. Multiple regression is generally much more powerful than simple linear regression. Compare our two examples.

With multiple regression, you must also consider the **significance** level of each **independent variable**. In the example above, the **significance** level of both **independent variables** is less than .001.

Phrasing Results That Are Significant

In our example, we obtained an R Square of .993 and a regression equation of WEIGHT' = 47.138 – 39.133(SEX) + 2.101(HEIGHT). The ANOVA resulted in F = 981.202 with 2 and 13 degrees of freedom. F is significant at the less than .001 level. Thus, we could state the following in a results section:

> A multiple linear regression was calculated to predict subjects' weight based on their height and sex. A significant regression equation was found ($F(2,13) = 981.202$, $p < .001$), with an R^2 of .993. Subjects' predicted weight is equal to 47.138 – 39.133(SEX) + 2.101(HEIGHT), where sex is coded as 1 = Male, 2 = Female, and height is measured in inches. Subjects increased 2.101 pounds for each inch of height, and males weighed 39.133 pounds more than females. Both sex and height were significant predictors.

The conclusion states the direction (increase), strength (.993), value (981.20), degrees of freedom (2,13), and **significance** level (< .001) of the regression. In addition, a statement of the equation itself is included. Because there are multiple **independent variables**, we have noted whether or not each is significant.

Phrasing Results That Are Not Significant

If the ANOVA does not find a significant relationship, the *Sig.* section of the output will be greater than .05, and the regression equation is not significant. A results section might include the following statement:

> A multiple linear regression was calculated predicting subjects' ACT scores based on their height and sex. The regression equation was not significant ($F(2,13) = 1.21$, $p > .05$) with an R^2 of .062. Neither height nor weight can be used to predict ACT scores.

Note that for results that are not significant, the ANOVA results and R^2 results are given, but the regression equation is not.

Practice Exercise

Use Practice Data Set 2 in Appendix B. Determine the prediction equation for predicting salary based on education, years of service, and sex. Which variables are significant predictors? If you believe that men were paid more than women, what would you conclude after conducting this analysis?