Covariance and Correlation

covariance—a measure of how one variable covaries with another variable

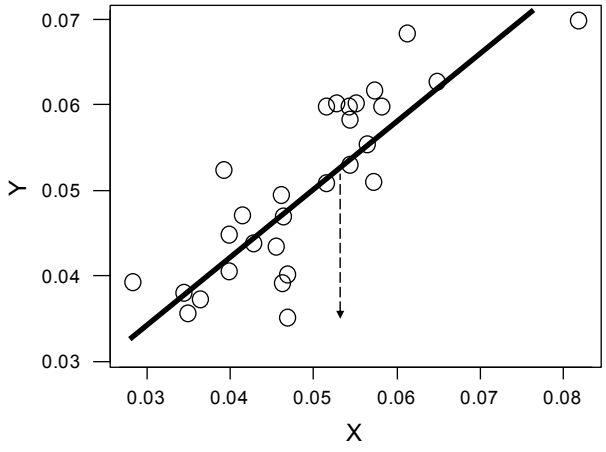
$$cov(x,y)=\frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N-1}$$

correlation—a standardized measure of how one variable covaries with another variable

$$r=\frac{cov(x,y)}{s_x s_y}=\frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{(N-1)s_x s_y}$$

where s_x = standard deviation of variable x
s_y = standard deviation of variable y

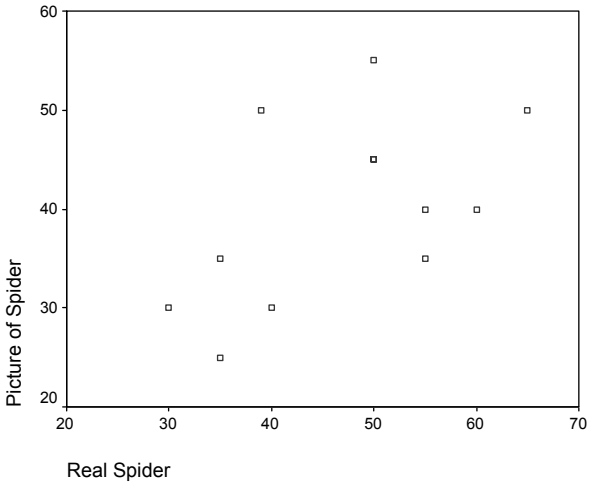
Visualizing through Scatterplots



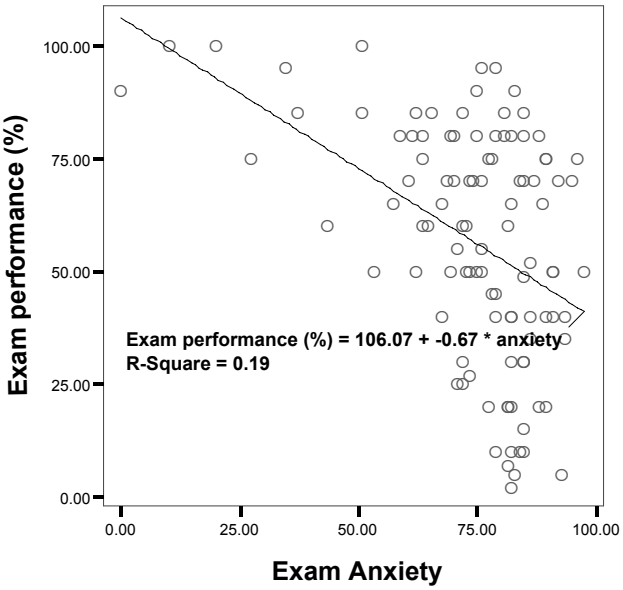
Generating Scatterplots in SPSS

Menu: Graphs→ scatter…→ simple
Menu: Interactive→scatterplots

Scatterplot of spider anxiety scores



Scatterplot with Regression Line



Bivariate Correlations

- Pearsons Product Moment– generates "r" value (statistic) whose value can range from -1 to +1
- Used with interval or ratio data
- Generates p-value to test the hypothesis whether relationship between 2 variables is significant

Bivariate Correlations

- Spearman's Rho– generates "r" value (statistic) whose value ranges from -1 to +1
- Use with non-parametric (non-normal) or ordinal data
- Generates p-value to test the hypothesis whether relationship between 2 variables is significant

Bivariate Correlations in SPSS

- Menu: Analyze→ Correlate
- Generate Pearsons Product Moment or Spearmans Rho

Correlations

		Picture of Spider	Real Spider
Picture of Spider	Pearson Correlation	1	R = .545
	Sig. (2-tailed)	.	.067
	N	12	12
Real Spider	Pearson Correlation	.545	1
	Sig. (2-tailed) p-value	.067	.
	N	12	12

Bivariate Correlations in SPSS

- Menu: Analyze→ Correlate
- Generate Pearsons Product Moment or Spearmans Rho

Correlations

			Picture of Spider	Real Spider
Spearman's rho	Picture of Spider	Correlation Coefficient	1.000	R = .490
		Sig. (2-tailed)	.	.106
		N	12	12
	Real Spider	Correlation Coefficient	.490	1.000
		Sig. (2-tailed) p-value	.106	.
		N	12	12

Coeffecient of Determination

- measures the amount of variance in one variable (dependent) that is explained by the other variable (independent)
- found by squaring r-value
- for example, if r=.545, then r^2=.297

Interpretation: 29.7 percent of the variation in y can be explained by the variation in x

Partial Correlations

- a correlation between two variables in which the effects of other variables are held constant

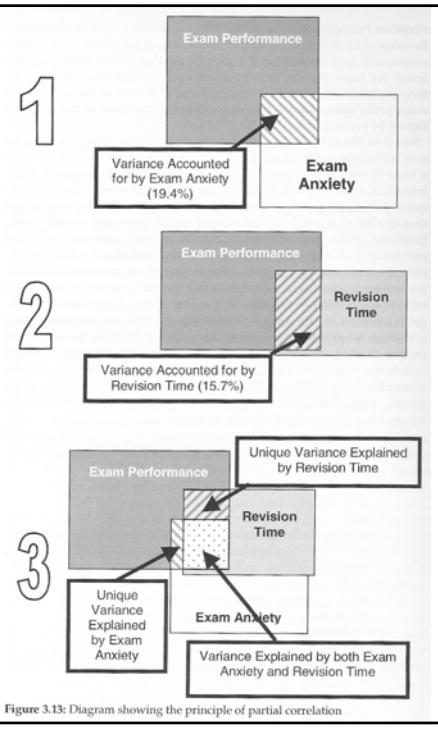


Figure 3.13: Diagram showing the principle of partial correlation

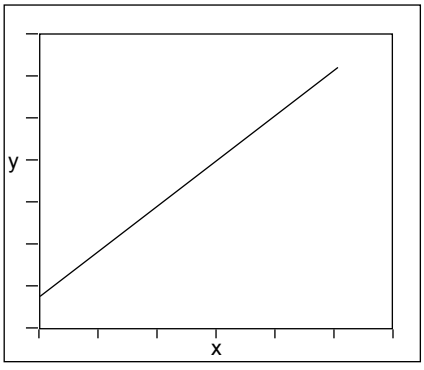
Source: Field, A. 2000. Discovering Statistics Using SPSS for Windows. Sage Publications. p. 98

Linear Regression

Estimation -- The General Linear Model

Formula for a straight line

$y = b_0 + b_1x$

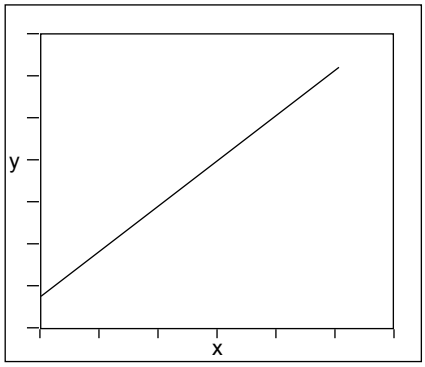


Estimation -- The General Linear Model

Formula for a straight line

dependent variable

$y = b_0 + b_1x$

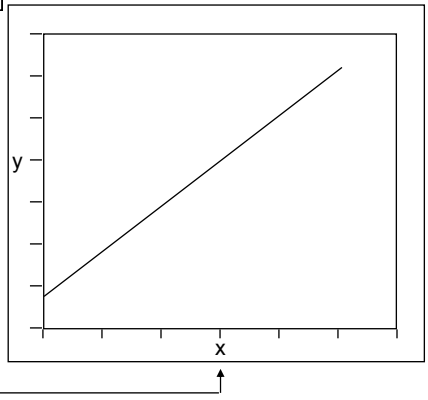


Estimation -- The General Linear Model

Formula for a straight line

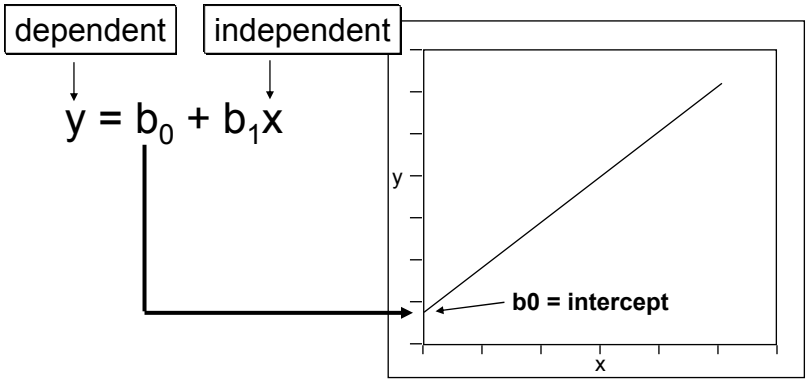
dependent variable | independent variable

$y = b_0 + b_1x$



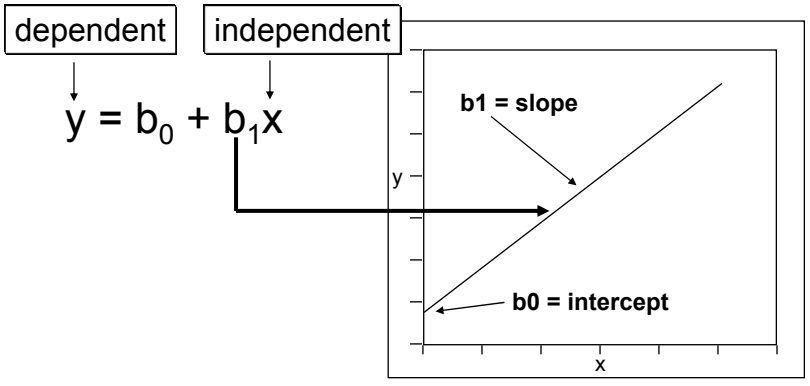
Estimation -- The General Linear Model

Formula for a straight line



Estimation -- The General Linear Model

Formula for a straight line

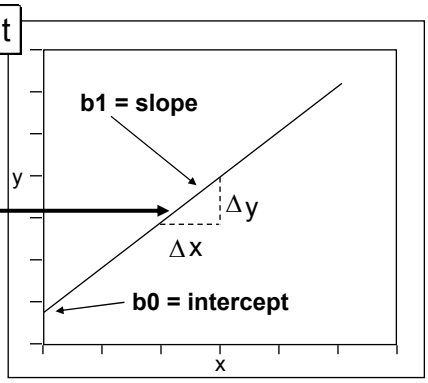


Estimation -- The General Linear Model

Formula for a straight line

dependent independent

$y = b_0 + b_1x$

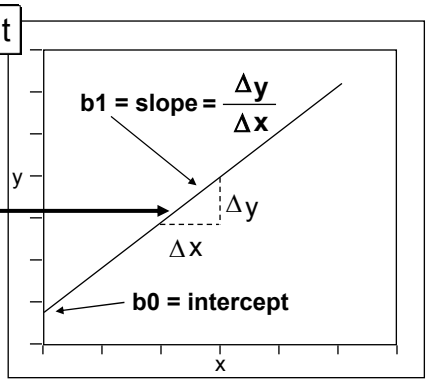


Estimation -- The General Linear Model

Formula for a straight line

dependent independent

$y = b_0 + b_1x$

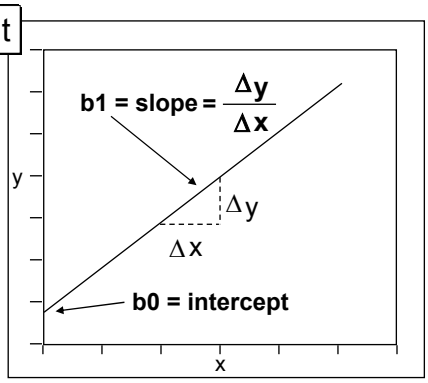


Estimation -- The General Linear Model

Formula for a straight line

dependent independent

$$y = b_0 + b_1 x$$



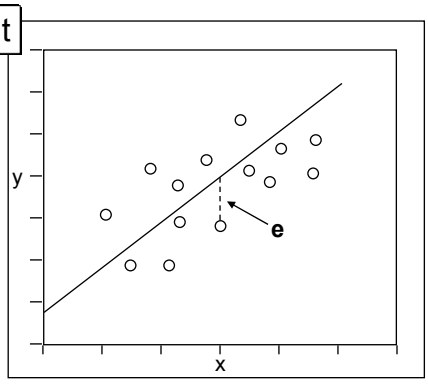
Estimation -- The General Linear Model

Formula for a straight line

dependent independent

$$y = b_0 + b_1 x + e$$

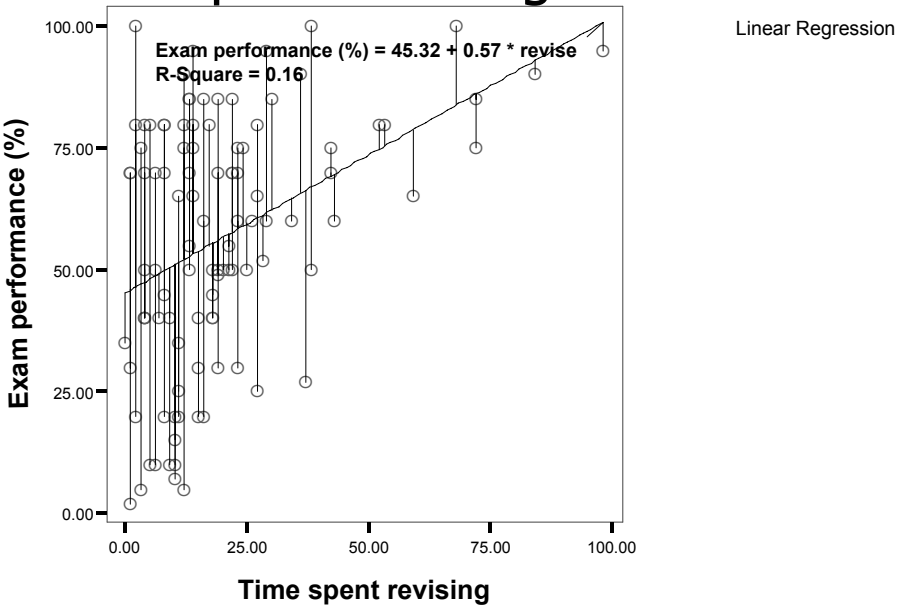
Error in measurement



Estimation- - The General Linear Model

- Least squares criterion
- Multiple regression -- adds more x's on right side
- Can estimate the variability of the points around the line

Scatterplot with Regression Line



Building a Predictive Model Using Simple Linear Regression in SPSS

- Found in SPSS: Regression→ Linear
- Need to specify independent and dependent variables

Example: Is record airplay time related to record sales?

What is strength and direction of the relationship?

How much variation is explained?

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.599[a]	.359	.355	64.78750

a. Predictors: (Constant), No. of Plays on Radio 1 per week

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	464862.8	1	464862.811	110.750	.000[a]
	Residual	831089.2	198	4197.420		
	Total	1295952	199			

a. Predictors: (Constant), No. of Plays on Radio 1 per week

b. Dependent Variable: Record Sales (thousands)

Is the predictive model statistically significant?

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	84.873	11.267		7.533	.000
	No. of Plays on Radio 1 per week	3.939	.374	.599	10.524	.000

a. Dependent Variable: Record Sales (thousands)

Is the independent variable statistically significant?