Multiple Regression -- Predicting a Quantitative Criterion Using 2 or More Quantitative (or Binary) Predictors

Application: To identify how multiple predictors contribute to the the linear equation relating the predictors to the criterion.

Research Hypothesis: There is the implicit hypothesis that the multivariate model significantly predicts the criterion. There might also be hypotheses about which predictors do and do not contribute to the multivariate model

Start with the Correlations!

It is a good idea to examine the correlations among the variables you plan to include in the multiple regression equation before performing the multiple regression analyses.

Correlations

		rating of fish quality - 1-10 scale	type of fish available	number of fish at store
rating of fish quality - 1-10 scale	Pearson Correlation	1	.595*	-.707*
	Sig. (2-tailed)	.	.041	.010
	N	12	12	12
type of fish available	Pearson Correlation	.595*	1	-.441
	Sig. (2-tailed)	.041	.	.151
	N	12	12	12
number of fish at store	Pearson Correlation	-.707*	-.441	1
	Sig. (2-tailed)	.010	.151	.
	N	12	12	12

*. Correlation is significant at the 0.05 level (2-tailed).

The criterion variable for this analysis will be the rating of fish quality. The two predictors will be the typw of fish available (a binary variable with the values 1 = freshwater only & 2 = fresh- and saltwater.

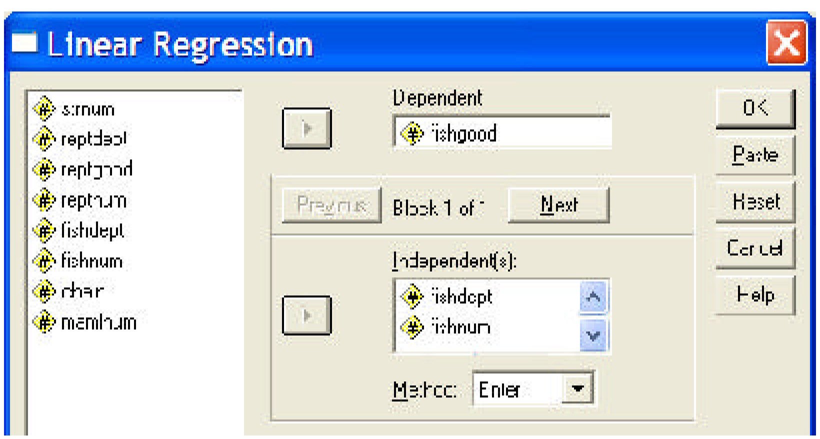
The correlation matrix shows that there is a significant positive correlation between ratings of fish quality and type of fish available in the stores. The interpretation of this is that the stores with fresh- and saltwater fish (with the higher code = 1) have a significantly higher mean fish quality ratings than stores with only freshwater fish (with the lower code = 0).

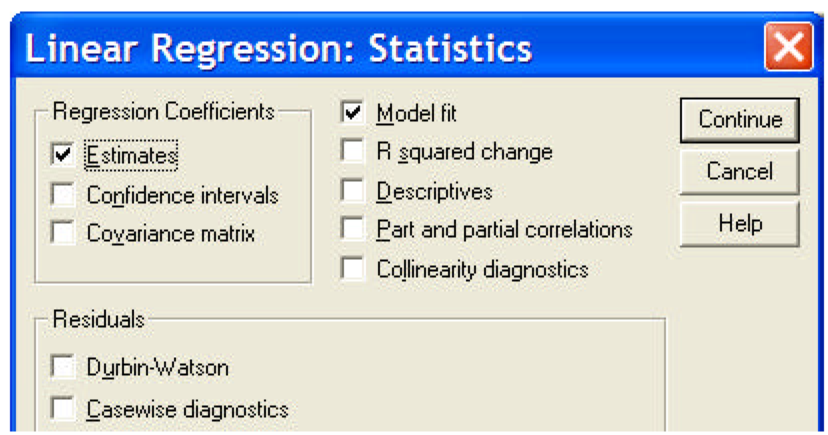
The matrix shows a significant negative correlation between ratings of fish quality and number of fish at the stores. The interpretation of this is that stores with more fish tended to have lower fish quality ratings.

On to the multiple regression...

Analyze → Regression → Linear

- Highlight the criterion variable and click the arrow to move it into the “Dependent” box
- Highlight the predictor variables and click the arrow to move them into the “Independent(s)” box
- Click “Statistics” -- in the **Linear Regression: Statistics** window be sure “Estimates” and “Model fit” are checked





Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.774[a]	.599	.510	1.535

a. Predictors: (Constant), number of fish at store, type of fish available

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	31.719	2	15.859	6.733	.016[a]
	Residual	21.198	9	2.355		
	Total	52.917	11			

a. Predictors: (Constant), number of fish at store, type of fish available

b. Dependent Variable: rating of fish quality - 1-10 scale

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	9.185	1.757		5.228	.001
	type of fish available	1.477	.987	.352	1.496	.169
	number of fish at stor	-.131	.056	-.552	-2.347	.044

a. Dependent Variable: rating of fish quality - 1-10 scale

The R is the multiple correlation coefficient.

R^2 tell the proportion of the criterion variable's variation is accounted for by the predictor

The ANOVA provides the significance test of the H0: that $R^2 = 0$ (or R=0).

The p-value < .001 tells us that the linear relationship beween these variables is statisically significant. We would not want to use a predictor that is not significantly related with the criterion.

Each "B" (usually written as "b") is the raw score multiple regression weight for that predictor . It tells the direction and extent of the change in the criterion for each 1-unit change in that predictor when holding the value of all other predictors constant.

The b for number of fish tells us that the rating of fish quality is expected to decrease by .131 for each additional fish, if the type of fish available is held constant. The p-value tells us that this predictor makes a significant contribution to the multiple regression model.

The b for type of fish available tells us that stores with fresh- and saltwater fish available (code = 1) have a mean fish quality that is 1.477 higher than stores with only freshwater fish available (code = 0) holding the number of fish constant. The p-value tells us that this mean difference is not significant and that this predictor is not contributing to the multiple regression model.

The standardized coefficients of "Beta" weights tellt the expected change in the criterion in Z-score units for each 1-Z-unit change in that predictor, holding the value of all other predictors constant. These Beta weights can be used to make a rough comparison of the relative contribution of the predictors to the model.

The constant (usually symbolized as "a") is the y-intercept. It tells the expected value of the criterion when the value of both predictors is 0

So, for these data the constant tells us that a store that has 0 fish and is has only freshwater fish (value = 0) it is expected to have a fish quality rating of 9.185 (though this makes little sense --how would you rate the quality of fish that aren't there?).

Computing predicted criterion scores for new cases

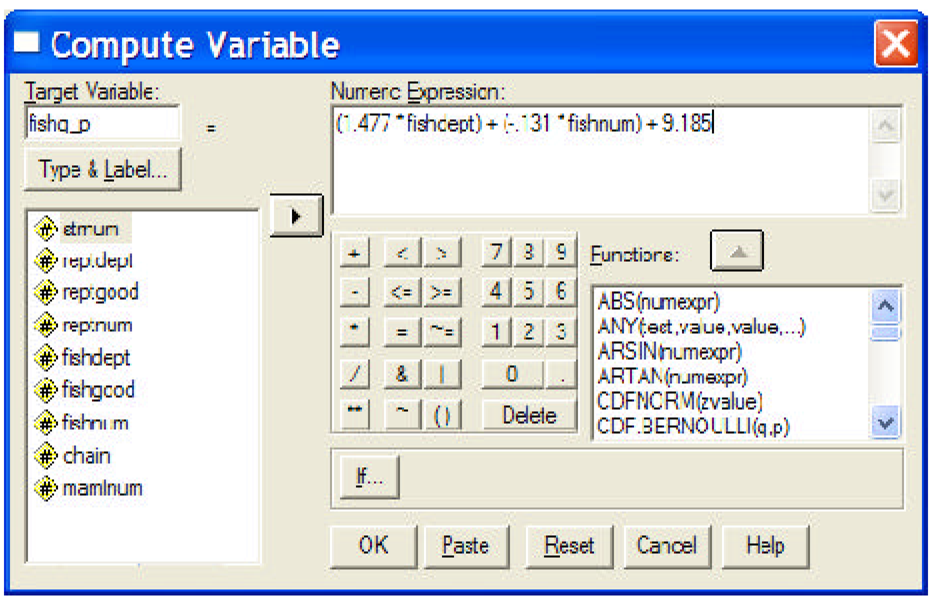
Once you have the multiple regression formula you can use it to predict criterion variable scores for new cases or participants for which you have only predictor variable scores. This is done much as with simple regression, but there will be multiple weighted predictors in te compute statement.

Please note that you must use all of the predictors included in the model, whether they contribute to the model or not -- it is not appropriate to include just a subset of the predictors. If you wish to eliminate predictors that do not contribute to the model you must rerun the model without those predictors and use the resulting regression weights.

Transform ➔ Compute

- Type the name of the predicted criterion variable in the "Target Variable:" box (1-8 characters)
- Type the regression formula into the "Numeric Expression" box

When you press "OK" a variable with the name you specified will be added into the rightmost column of the data editor for each case.



Write-up

Correlation and multiple regression analyses were used to explore the relationship between fish number, type of fish available and fish quality. Table 1 shows the univariate statistics for these variables.

Table 2 shows the correlation and regression results. Fish quality is significantly negatively correlated with fish qualtiy, indicating that stores with more fish tend to have fish of lower overall quality. Type of fish department was significantly postively correlated with fish quality, revealing that stores with both types of fish tended to have fish of higher overall quality than stores with only freshwater fish.

The multiple regression results show that there is a significant negative relationship between number of fish and fish quality after taking type of fish available into account, however, type of fish available at a store does not contribute to this multivariate model.

Table 1.
Univariate statistics for criterion and predictor variables (N=12)

Variable	Univariate Statistics	
Fish Quality	M = 6.58	SD = 2.10
Number of Fish	M = 25.59	SD = 9. 27
Type of Fish Available	Freshwater Fish	n = 6 (50%)
	Freshwater & Saltwater Fish	n = 6 (50%)

Table 2.
Correlations and multiple regression weights

Variable	r (p)	b (p)	ß
Number of Fish	-.707 (.01)	-0.131 (.04)	-.552
Type of Fish Available	.595 (.04)	1.447 (.17)	.352
constant		9.158	