Graduate Statistics

Multivariate Correlational Analysis

What We Will Cover in This Section

- Introduction
- Multiple Regression
- Canonical Correlation

Research Problem

A group of students conducted a validation study of the Activity Vector Analysis (AVA). The traditional interpretation of this assessment tool was that people scoring higher on the Sociability scale would be more optimistic than those scoring lower on the scale.

The AVA measures five scales Aggression (V-1), Sociability (V-2), Calmness (V-3), Conformity (V-4), and Conscious Restraint (V-5). One of the criterion measures was the AASQ, a standardized measure of Optimism.

Correlational Tests

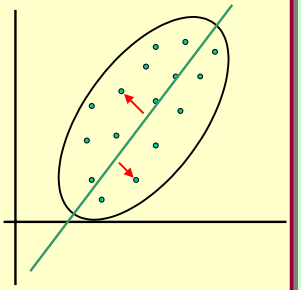
Test	Predictors	Criterion
Pearson r	One Continuous	One Continuous
Multiple R	Multiple Continuous	One Continuous
Canonical R	Multiple Continuous	Multiple Continuous

General Model

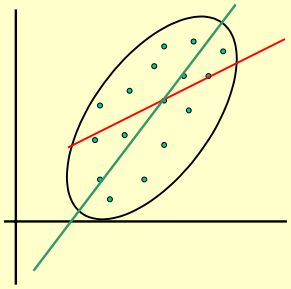
Subjects	Variables				
	x_1	x_2	x_3	…	x_k
S_1					
S_2					
S_3					
⋮					
S_n					

Terminology: Residual

This is the error in prediction and represents the amount of variance not accounted for in prediction.



Linear Equations: Least Squares Criterion



$$Y = bX + a$$

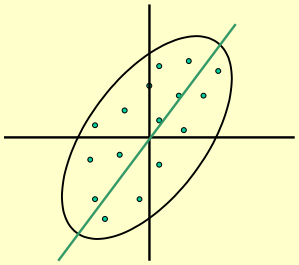
Terminology: Regression Coefficient (B)

This is the weight given to each predictor in the regression equation when using raw scores.

Terminology: Beta Weight (ß)

This is the weight given to each predictor in the regression equation when using z-scores (standard scores).

Linear Equations: Z-scores



$$\hat{Z}_Y = r \times Z_X$$

$$\hat{Z}_Y = \beta \times Z_X$$

Multiple Regression Equations

$$\hat{Y} = B_1X_1 + B_2X_2 + \ldots + B_kX_k + \Lambda$$

$$\hat{Z}_Y = \beta_1 Z_{X_1} + \beta_2 Z_{X_2} + \ldots + \beta_k Z_{X_k}$$

Multiple Correlation Coefficient

- **Indicated by R.**
- **R^2 is equivalent to the *Coefficient of Determination*.**
- **1 - R^2 is equivalent to the *Coefficient of Alienation*.**

EXTRORDINARILY
COOL FACT

The the bivariate correlation (r) between Y' and Y equals R.

Significance Test

- Use *F-test* to assess statistical significance.

$$F = \frac{\text{Mean Treatment Variability}}{\text{Mean Error Variability}}$$

$$F_{(k),(N-k-1)} = \frac{R^2 / k}{(1-R^2)/(N-k-1)}$$

Issue #1, Multicollinearity

- A problem that arises when two or more of the predictors are highly correlated.
  - Severely limits the size of R.
  - Individual effects are confounded because of overlapping information.
  - Causes unstable regression coefficients.

Finding Multicollinearity

1. Examine the correlation matrix and identify highly correlated variables (r > .80)
2. Compute *TOLERANCE*.
   - A statistic that indicates multicollinearity among the variables.
   - Ranges from 0 to 1 with higher indicating lower intercorrelations.

Dealing With Multicollinearity

- Delete one of the variables.
- Combine the variables.

Issue #2, Selecting Variables

- Avoid Sweeney Stew Empiricism.
- Be parsimonious in selecting variables for inclusion in the analysis.

Multiple Regression Example

A group of students conducted a validation study of the Activity Vector Analysis (AVA). The traditional interpretation of this assessment tool was that people scoring higher on the Sociability scale would be more optimistic than those scoring lower on the scale.

The AVA measures five scales Aggression (V-1), Sociability (V-2), Calmness (V-3), Conformity (V-4), and Conscious Restraint (V-5). One of the criterion measures was the AASQ, a standardized measure of Optimism.

	AASQ	Assert	Sociab	Calm.	Conform	Consc. Restraint
AASQ	1.00					
Assertiveness	.038	1.00				
Sociability	.124	.459**	1.00			
Calmness	.128	.207*	.401**	1.00		
Conformity	-.344**	.528**	.240**	.180*	1.00	
Conscious Restraint	.185*	.683**	.491**	.493**	.369**	1.00

* p < .05, ** p <.01

Multiple Regression Techniques

Standard Multiple Regression

- All of the variables are entered at once.
- Provides an indication of the amount of variability that each variable contributes to the criterion variable.
- Problem.
  - It capitalizes on chance relationships.

Simultaneous Regression

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.468[a]	.219	.180	5.7323	.219	5.512	5	98	.000

a. Predictors: (Constant), Conscious Restraint, C score, Conformity, C score, Calmness, C score, Sociability, C score, Assertiveness, C score

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	48.918	4.692		10.427	.000		
	Assertiveness, C score	8.851E-02	.108	.116	.817	.416	.402	2.485
	Sociability, C score	1.094E-02	.071	.015	.155	.877	.740	1.351
	Calmness, C score	5.567E-02	.068	.084	.813	.418	.746	1.341
	Conformity, C score	-.283	.062	-.475	-4.572	.000	.739	1.354
	Conscious Restraint, C score	.153	.097	.211	1.572	.119	.442	2.263

a. Dependent Variable: AASQ Total

Stepwise Multiple Regression

- Forward and Stepwise.
  - Predictors are entered one at a time starting with the variable that has the highest correlation with the criterion variable.
  - After each predictor is entered statistical tests are conducted to determine if the the next variable will significantly enhance R or R^2.
- Backward
  - All predictors are entered.
  - Significance tests are conducted for each variable as if it were the last one entered.
  - If the variable does not contribute a significant amount to R it is dropped.

Stepwise Regression

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.344[a]	.118	.110	5.9714	.118	13.707	1	102	.000
2	.458[b]	.209	.194	5.6829	.091	11.618	1	101	.001

a. Predictors: (Constant), Conformity, C score

b. Predictors: (Constant), Conformity, C score, Conscious Restraint, C score

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	58.638	3.169		18.501	.000		
	Conformity, C score	-.206	.056	-.344	-3.702	.000	1.000	1.000
2	(Constant)	51.869	3.611		14.363	.000		
	Conformity, C score	-.262	.055	-.439	-4.732	.000	.911	1.098
	Conscious Restraint, C score	.229	.067	.316	3.409	.001	.911	1.098

a. Dependent Variable: AASQ Total

Cross-validation

- **FACT. Multiple regression maximizes the least squares criterion for the sample on which the analysis is being performed.**
- **SHRINKAGE. The reduction in R when an equation developed in one sample is used in another sample.**
- **WHY? Different samples will yield different results because of sampling error.**

How to Cross-validate

- **Replicate the study with another random sample from the same population.**
- **Split the original sample into two sub-samples and develop an equation for each sub-sample.**

Canonical Correlation

Research Problem

- **Predictors.**
  - Seat location.
  - Grade point average.
  - Commuter status.
  - Working hours.
  - Verbal skills.
  - Quantitative skills.
- **Criterion.**
  - Class grade.
  - Partying.
  - Class participation.
  - Extracurricular activities.
  - Class attendance.

Variables, Variates, and Variate Pairs

Variate 1

GPA+Verbal+ Quantitative ⟺ Grade + Attendance

Variate 2

Commuter Status+ Working hours ⟺ Attendance + Extracurricular

What You Learn

1. What are the variates?
2. How many variate pairs are there?
3. How to interpret the variates?
4. What is the correlation between the variate pairs?
5. Are the canonical variate scores interpretable?

Issues

1. Number of cases required.
2. Assumption of normality.
3. Missing data.
4. Absence of outliers.
5. Absence of multicollinearity.

THE END