

# Multivariate Analysis of Variance



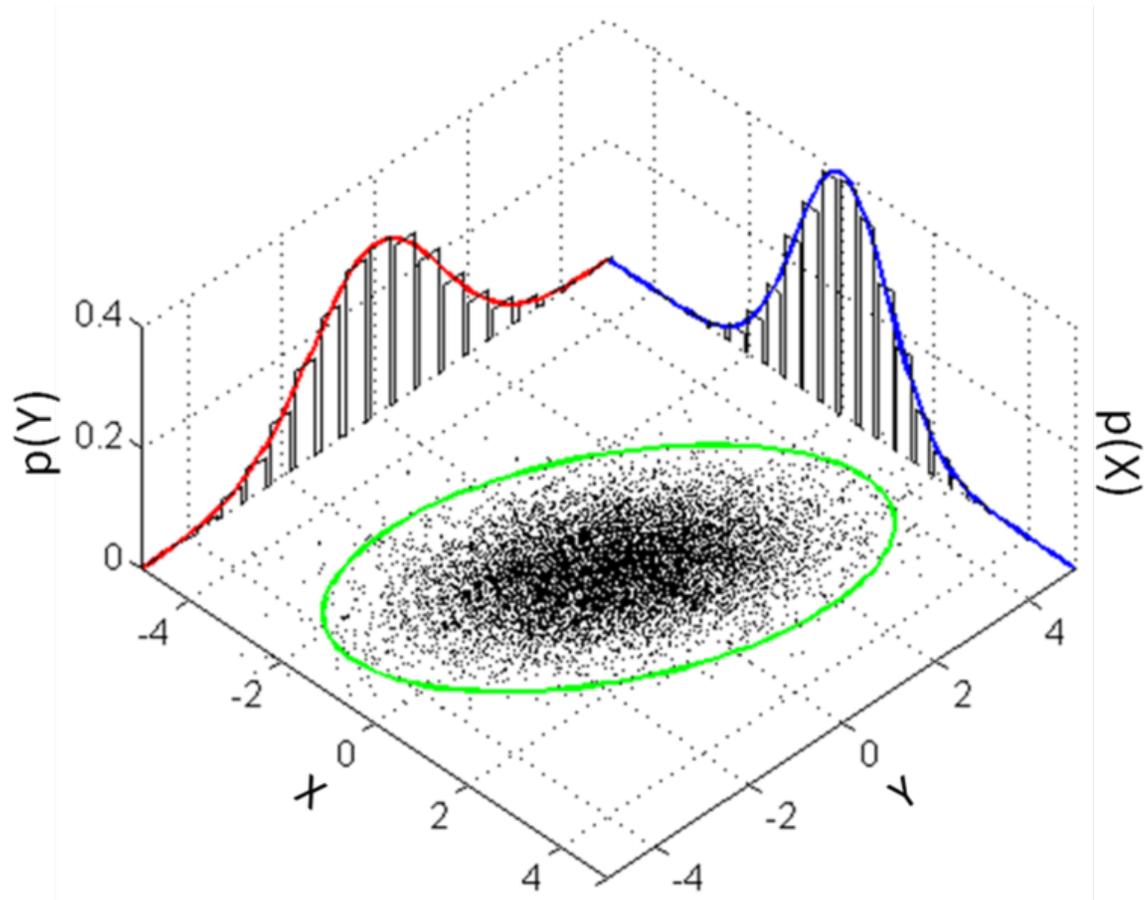
Introduction to Statistics  
Carl von Ossietzky Universität Oldenburg  
Fakultät III - Sprach- und Kulturwissenschaften

## Introduction

- Multivariate analysis of variance (MANOVA) is simply an ANOVA with several dependent variables.
- Univariate: one dependent variable, multivariate: several dependent variables.
- Example: Mehnert(1998) investigated the effect of planning time on L2 speech performance.
- Independent variable: planning time (several levels).
- Dependent variables: four different measures of speech performance.
- A significant effect (Pillais=0.92,  $F=2.75$ ,  $p < 0.005$ ) was found for planning time on the four dependent variables.
- MANOVA has greater power to detect an effect than a series of ANOVA's, since MANOVA can detect whether groups differ along a *combination* of variables.

## Assumptions

- Independence  
Observations should be statistically independent.
- Random sampling  
Data should be randomly sampled from the population of interest and measured at an interval level.
- Multivariate normality  
The dependent variables (collectively) have multivariate normality within groups.
- Homogeneity of covariance matrices  
The variances in each groups for each of the dependent variables are roughly equal. The correlation between any pair of dependent variables is the same in all groups. Test whether the covariance matrices of the different groups in the analysis are equal.



Multivariate normality. Picture taken from *Wikipedia: Multivariate normal distribution*.

## Multivariate normality

- Check whether the dependent variables (collectively) have multivariate normality within groups.
- To test for multivariate normality the Mahalanobis distance is used. This test identifies multivariate outliers which influence normality.
- In SPSS: add a new column with arbitrary numbers in the table. For example a column *Id* containing index numbers.
- Go to Analyze, Regression, Linear.
- Move *Id* under 'Dependent', and your dependent variables (i.e. in de MANOVA analysis they are dependent) under 'Independent(s)'.
- Click on Save, and choose Mahalanobis. Click on Continue and on OK.
- SPSS has added an extra column *MAH\_1* in the table.
- The Mahalanobis Distance score for each subject is considered an outlier if it exceeds a "critical value".

## Multivariate normality

- The degrees of freedom for this test is equal to the number of variables under investigation. If there are  $n$  dependent variables, the degrees of freedom is  $n$ .
- The probability level you set for this test is  $p < .001$ .
- The critical chi-square values for 2 to 10 degrees of freedom at a critical alpha of .001:

$df$	Critical value
2	13.82
3	16.27
4	18.47
5	20.52
6	22.46
7	24.32
8	26.13
9	27.88
10	29.59

## Multivariate normality

- For other chi-square values go to <http://www.vassarstats.net/> and choose Distributions, Chi-Square Distributions and enter the degrees of freedom.
- The table at the right *Critical values of chi-square for  $df=n$*  gives you the chi-square values for several alpha's.
- Check which of the Mahalanobis Distance scores (in column added by SPSS) is higher than the critical value. Those scores are the outliers!

## Homogeneity of covariance matrices

- Preliminary check: Levene's test per dependent variable.
- Homogeneity of covariance matrices is tested with Box's test, which tests the hypothesis that the covariance matrices of the dependent variables are equal across levels of the independent variable.
- If Box's test is significant, you have violated an assumption of MANOVA.
- This is not much of a problem if you have equal cell sizes and large  $n$ . It is a much bigger issue with small sample sizes and/or unequal cell sizes.
- But what exactly are covariance matrices?

## Covariance matrix

- Covariance: a measure of how much two random variables change together.
- If the larger values of one variable mainly correspond with the larger values of the other variable, and the same holds for the smaller values, i.e., the variables tend to show similar behavior, the covariance is positive.
- A correlation coefficient is a covariance scaled between -1 and +1.
- Assume three dependent variables  $V1$ ,  $V2$  and  $V3$ . We calculate covariance between the variables:

	$V1$	$V2$	$V3$
$V1$	78	21	33
$V2$	21	91	12
$V3$	33	12	85

- When a covariance is 0, there is no linear relationship between the variables.
- The diagonal variables are the variances of each variable. A covariance of a variable with itself is simply the variance.

## MANOVA on SPSS

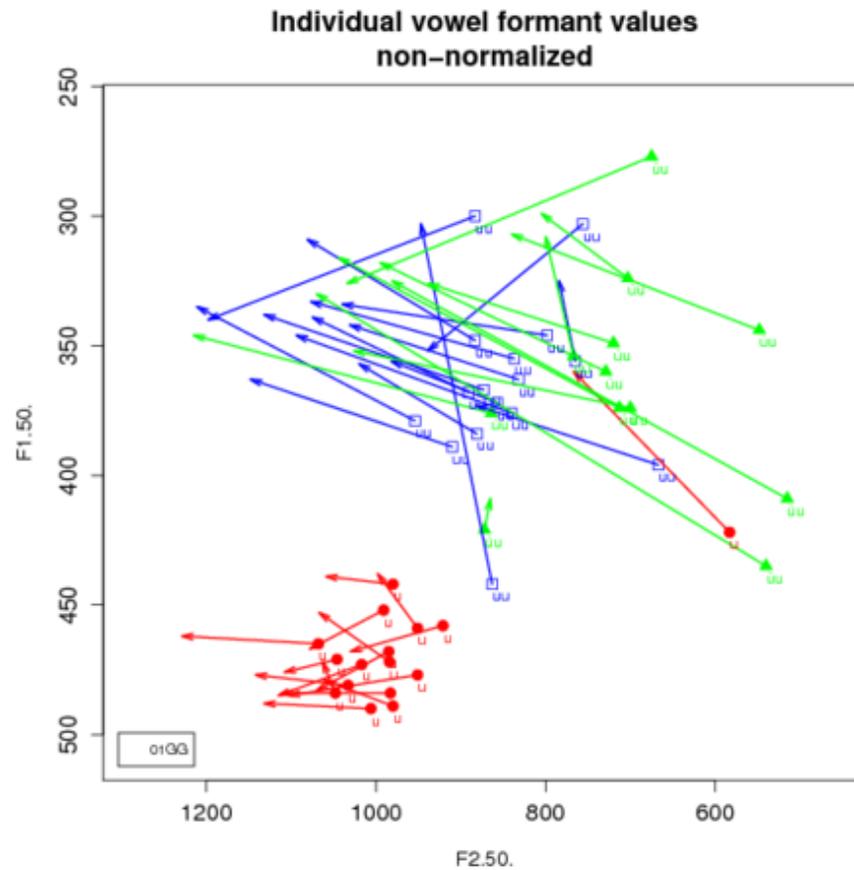
- Go to Analyze, General Linear Model, Multivariate.
- Move the dependent variables under 'Dependent variables', and the independent variable(s) under 'Fixed factor(s)'.
- Under Post Hoc move all factors into 'Post Hoc Tests for'. Select 'Bonferroni'.
- Under Options, move all factors and interactions into the 'Display Means for window'. Select 'descriptive statistics', 'estimates of effect size' and 'homogeneity tests'. Set the confidence level to .05 and click on Continue. Click on OK.

## Example

- Example from DFG project *Lautliche und prosodische Variation im Saterland*. The project is under supervision of Prof. Dr. Jörg Peters and carried out by Heike Schoormann and Wilbert Heeringa (Institut für Germanistik).
- F1, F2 and vowel duration are measured in three Saterfrisian words pronounced by one native-speaker:

German	Saterfrisian	Number
voll	ful	7
faul	fuul	5
viel	fúul	11

- Do F1, F2 and vowel duration distinguish the vowels in the three words?



F1 and F2 values of three vowel groups: u (red), uu (blue), úu (green). Formants were measured at 50% and 80% time points in the vowel intervals. Arrows run from the 50% values to the 80% values.

## Testing assumptions

- Independence  
Observations are independent.
- Random sampling  
Data is randomly sampled and measured at an interval level.
- Multivariate normality  
The dependent variables (collectively) have multivariate normality within groups.
- We have three dependent variables, therefore  $df=3$ . For  $\alpha=0.0001$  we find critical chi-square = 16.27.
- Our highest Mahalanobis distance score is 6.49, therefore, we do not have outliers.

## Testing assumptions

- Homogeneity of covariance matrices  
Test whether the population covariance matrices of the different groups in the analysis are equal.
- First we check the results of Levene's test:

**Levene's Test of Equality of Error Variances<sup>a</sup>**

	F	df1	df2	Sig.
Dur V	,001	2	20	,999
Formant1 50	3,260	2	20	,060
Formant2 50	,066	2	20	,937

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Lexeme

- We find  $p > 0.05$  for the three dependent variables, the groups have the same variances for each of the dependent variables.

## Testing assumptions

- Results of Box's test:

**Box's Test of  
Equality of  
Covariance  
Matrices<sup>a</sup>**

Box's M	20,663
F	1,261
df1	12
df2	796,139
Sig.	,237

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design:  
Intercept +  
Lexeme

- We find  $p > 0.05$ , the covariance matrices of the three groups are the same.

## Multivariate tests

**Multivariate Tests<sup>a</sup>**

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	,996	1683,349 <sup>b</sup>	3,000	18,000	,000	,996
	Wilks' Lambda	,004	1683,349 <sup>b</sup>	3,000	18,000	,000	,996
	Hotelling's Trace	280,558	1683,349 <sup>b</sup>	3,000	18,000	,000	,996
	Roy's Largest Root	280,558	1683,349 <sup>b</sup>	3,000	18,000	,000	,996
Lexeme	Pillai's Trace	1,083	7,486	6,000	38,000	,000	,542
	Wilks' Lambda	,167	8,681 <sup>b</sup>	6,000	36,000	,000	,591
	Hotelling's Trace	3,488	9,882	6,000	34,000	,000	,636
	Roy's Largest Root	2,986	18,909 <sup>c</sup>	3,000	19,000	,000	,749

a. Design: Intercept + Lexeme

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

We consider Pillai's test statistic, which is robust when sample sizes are equal.

## ANOVA summary table for dependent variables

**Tests of Between-Subjects Effects**

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Dur V	81133,243 <sup>a</sup>	2	40566,622	28,343	,000	,739
	Formant1 50	37028,998 <sup>b</sup>	2	18514,499	10,198	,001	,505
	Formant2 50	216230,28 <sup>c</sup>	2	108115,138	6,785	,006	,404
Intercept	Dur V	428948,517	1	428948,517	299,692	,000	,937
	Formant1 50	3270079,49	1	3270079,49	1801,220	,000	,989
	Formant2 50	13532745,3	1	13532745,3	849,237	,000	,977
Lexeme	Dur V	81133,243	2	40566,622	28,343	,000	,739
	Formant1 50	37028,998	2	18514,499	10,198	,001	,505
	Formant2 50	216230,277	2	108115,138	6,785	,006	,404
Error	Dur V	28625,974	20	1431,299			
	Formant1 50	36309,610	20	1815,481			
	Formant2 50	318703,636	20	15935,182			
Total	Dur V	686462,000	23				
	Formant1 50	3664284,00	23				
	Formant2 50	14808767,0	23				
Corrected Total	Dur V	109759,217	22				
	Formant1 50	73338,609	22				
	Formant2 50	534933,913	22				

a. R Squared = ,739 (Adjusted R Squared = ,713)

b. R Squared = ,505 (Adjusted R Squared = ,455)

c. R Squared = ,404 (Adjusted R Squared = ,345)

## Multiple comparisons

### Multiple Comparisons

Bonferroni

Dependent Variable	(I) Lexeme	(J) Lexeme	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval		
						Lower Bound	Upper Bound	
Dur V	u	uu	-39,29	22,152	,274	-97,16	18,59	
		úu	-131,92*	18,292	,000	-179,71	-84,13	
	uu	u	39,29	22,152	,274	-18,59	97,16	
		úu	-92,64*	20,405	,001	-145,95	-39,33	
	úu	u	131,92*	18,292	,000	84,13	179,71	
		uu	92,64*	20,405	,001	39,33	145,95	
	Formant1 50	u	uu	90,71*	24,949	,005	25,53	155,90
			úu	85,44*	20,601	,001	31,62	139,26
uu		u	-90,71*	24,949	,005	-155,90	-25,53	
		úu	-5,27	22,981	1,000	-65,31	54,77	
úu		u	-85,44*	20,601	,001	-139,26	-31,62	
		uu	5,27	22,981	1,000	-54,77	65,31	
Formant2 50		u	uu	108,00	73,915	,479	-85,11	301,11
			úu	223,18*	61,034	,005	63,73	382,64
	uu	u	-108,00	73,915	,479	-301,11	85,11	
		úu	115,18	68,086	,319	-62,70	293,06	
	úu	u	-223,18*	61,034	,005	-382,64	-63,73	
		uu	-115,18	68,086	,319	-293,06	62,70	

Based on observed means.

The error term is Mean Square(Error) = 15935,182.

\*. The mean difference is significant at the ,05 level.

## Linear Discriminant Analysis

- Linear Discriminant Analysis (LDA) can be used after MANOVA to see how the dependent variables discriminate the groups.
- In MANOVA we predict a set of outcome measures from a grouping variable, whereas in LDA we predict a grouping variable from a set of outcome measures.
- Looks a little like logistic regression analysis, but we may find more than just two groups. Logistic regression analysis, however, is more robust.
- LDA identifies variates that discriminate groups of cases. **Variates**: combinations of dependent variables.
- **Coefficients of linear discriminants** tell us how dependent variables contribute to the variates.
- The higher a coefficient for a dependent variable, the more important this variable is. Coefficients are either positive or negative.

## Wilk's Lambda

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	,167	34,002	6	,000
2	,666	7,731	2	,021

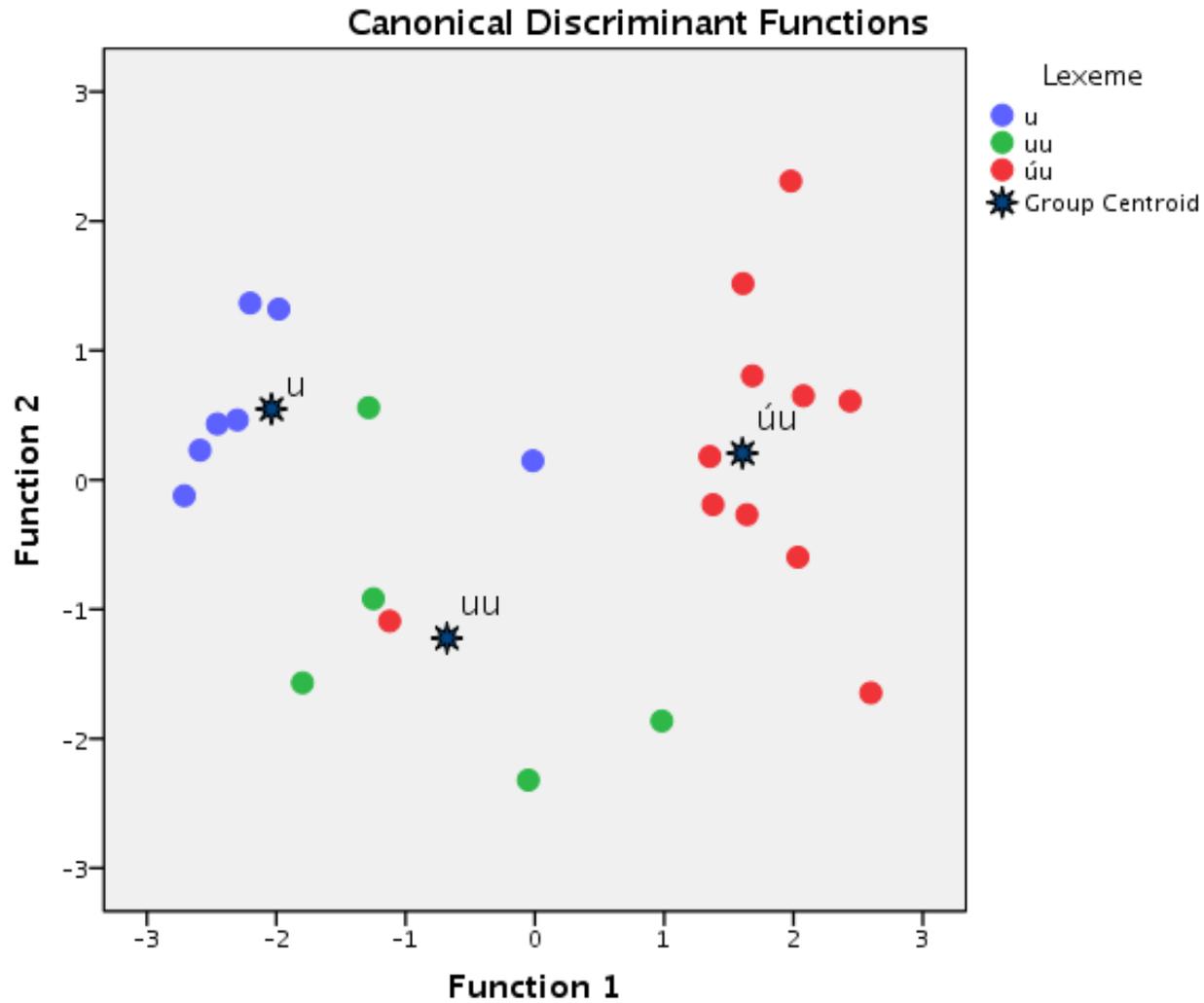
LDA identifies variates. If Wilk's Lambda is less than .05 then the variate is significantly discriminating the groups. '1 through 2' gives the significance of both variates, and '2' gives the significance after the first variate has been removed.

## Standardized Canonical Discriminant Function Coefficients

Standardized Canonical  
Discriminant Function  
Coefficients

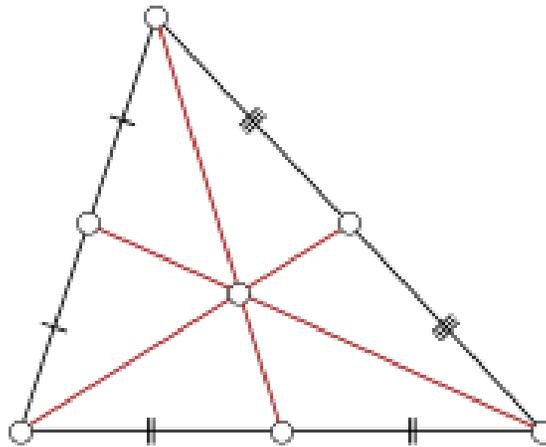
	Function	
	1	2
Dur V	,994	,697
Formant1 50	-,206	,995
Formant2 50	,135	,439

How do dependent variables contribute to the variates? High scores indicate that a dependent variable is important for a variate.



Linear Discriminant plot for the Saterfrisian fúul/fuul/ful triple.

## What are centroids?



The centroid or geometric center of a two-dimensional region is, informally, the point at which a cardboard cut-out of the region could be perfectly balanced on the tip of a pencil. Formally, the centroid of a plane figure or two-dimensional shape is the arithmetic mean ("average") position of all the points in the shape. Picture taken from *Wikipedia: Centroid*.

## Functions at Groups Centroids

Functions at Group  
Centroids

Lexeme	Function	
	1	2
u	-2,037	,548
uu	-,681	-1,222
úu	1,606	,207

Unstandardized canonical  
discriminant functions  
evaluated at group means

Table showing the coordinates of the centroids. Which groups are discriminated by a variate? Groups with values opposite in sign are being discriminated by that variate.

## Classification results

**Classification Results<sup>a</sup>**

			Predicted Group Membership			Total
			u	uu	úu	
Original	Count	Lexeme				
		u	6	1	0	7
		uu	1	4	0	5
		úu	0	1	10	11
	%	u	85,7	14,3	,0	100,0
		uu	20,0	80,0	,0	100,0
		úu	,0	9,1	90,9	100,0

a. 87,0% of original grouped cases correctly classified.

Classification table.  $((6 + 4 + 10) / (7 + 5 + 11)) \times 100 = 87.0\%$  of the vowels has been classified correctly.

## Classification results

**Classification Results<sup>a</sup>**

			Predicted Group Membership			Total
			u	uu	úu	
Original	Count	u	6	1	0	7
		uu	2	2	1	5
		úu	0	1	10	11
	%	u	85,7	14,3	,0	100,0
		uu	40,0	40,0	20,0	100,0
		úu	,0	9,1	90,9	100,0

a. 78,3% of original grouped cases correctly classified.

How well does vowel duration predict the three groups?

## Classification results

**Classification Results<sup>a</sup>**

			Predicted Group Membership			Total
			u	uu	úu	
Original	Count	u	7	0	0	7
		uu	1	2	2	5
		úu	2	5	4	11
%		u	100,0	,0	,0	100,0
		uu	20,0	40,0	40,0	100,0
		úu	18,2	45,5	36,4	100,0

a. 56,5% of original grouped cases correctly classified.

How well do F1 frequencies predict the three groups?

## Classification results

**Classification Results<sup>a</sup>**

			Predicted Group Membership			Total
			u	uu	úu	
Original	Count	u	6	0	1	7
		uu	2	2	1	5
		úu	2	1	8	11
	%	u	85,7	,0	14,3	100,0
		uu	40,0	40,0	20,0	100,0
		úu	18,2	9,1	72,7	100,0

a. 69,6% of original grouped cases correctly classified.

How well do F2 frequencies predict the three groups?