



# PRO2514 - Pesquisa Quantitativa em Gestão de Operações

## Análise Discriminante

Prof. Dr. Renato de Oliveira Moraes



# Análise Discriminante

- Quais variáveis causam maior divergência ou distinguem mais os grupos de indivíduos?
  - **Grupos de indivíduos:**
    - **Sexo**
    - **Idade**
    - **Classe social**
    - **Renda**
    - **Escolaridade**
    - **Clusters (extraídos - análise de Clusters)**
    - **Outras variáveis de agrupamento.**

Variáveis  
Não Métricas  
(nominais)



# Análise Discriminante

Análise de regressão  $\Rightarrow Y = f(X_1, X_2, \dots, X_n)$

tanto X quanto Y têm que ser variáveis métricas

Y = variável dependente

X<sub>1</sub>, X<sub>2</sub>, ..., X<sub>n</sub> = variáveis independentes

Análise discriminante  $\Rightarrow$

“Y” tem que ser nominal (não métrica) e

todas as variáveis “X” tem que ser métricas (ou assumidas como tal).



# Objetivos da Análise Discriminante

- **Verificar se existem diferenças significativas entre os grupos, considerando-se as variáveis independentes;**
- **Determinar quais são as variáveis independentes que mais contribuem para que os grupos sejam diferentes;**
- **Avaliar a precisão da classificação.**



# Estatísticas de classificação (*Classification Statistics*)

- **Classification Function Coeficients (coeficientes da função de classificação):** As Funções de Fischer são usadas para determinar a que grupo uma observação deve ser alocada. Para cada grupo (valor) da variável dependente (nominal) haverá uma função
- **Classification Results (resultados da classificação):** É apresentado um quadro que contabiliza a quantidade de indivíduos havia em cada grupo antes e depois da classificação pela análise discriminante, e o percentual de casos classificados corretamente.



# Exemplo 1

## **Variável dependente com 2 categorias**

**C12 - Especificação da compra**

0: não usa análise valor

1: usa análise de valor para cada compra

## **Variáveis independentes**

**C2: Delivery Speed**

**C3: Price Level**

**C4: Price Flexibility**

**C5: Manufacturer Image**

**C6: Service**

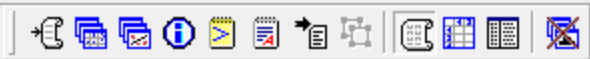
**C7: Salesforce Image**

**C8: Product Quality**



- Basic Statistics
- Regression
- ANOVA**
- DOE
- Control Charts
- Quality Tools
- Reliability/Survival
- Multivariate
- Time Series
- Tables
- Nonparametrics
- EDA
- Power and Sample Size

- One-Way...
- One-Way (Unstacked)...
- Two-Way...
- Analysis of Means...
- Balanced ANOVA...
- General Linear Model...
- Fully Nested ANOVA...
- Balanced MANOVA...
- General MANOVA...
- Test for Equal Variances...
- Interval Plot...
- Main Effects Plot...
- Interactions Plot...



Session

Level	N	Mean
0	40	2,500
1	60	4,192

Pooled StDev = 1,00

**One-Way Analysis of Variance**

Response: Delivery Speed

Factor: Specification Buying

Store residuals

Store fits

Confidence level: 95,0

Select Comparisons... Graphs...

Help OK Cancel

Worksheet 1 \*\*\*

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
	Id	Delivery Speed	Price Level	Price Flexibility	Manufacturer Image	Service	Salesforce Image	Product Quality	Firm Size	Usage Level	Satisfaction Level	Specification Buying
1	1	4	1	7	5	2	2	5	0	32	4	1
2	2	2	3	6	7	3	4	8	1	43	4	0
3	3	3	5	6	6	4	3	8	1	48	5	0
4	4	3	1	7	6	2	2	8	1	32	4	0
5	5	6	1	10	8	3	5	5	0	68	7	1

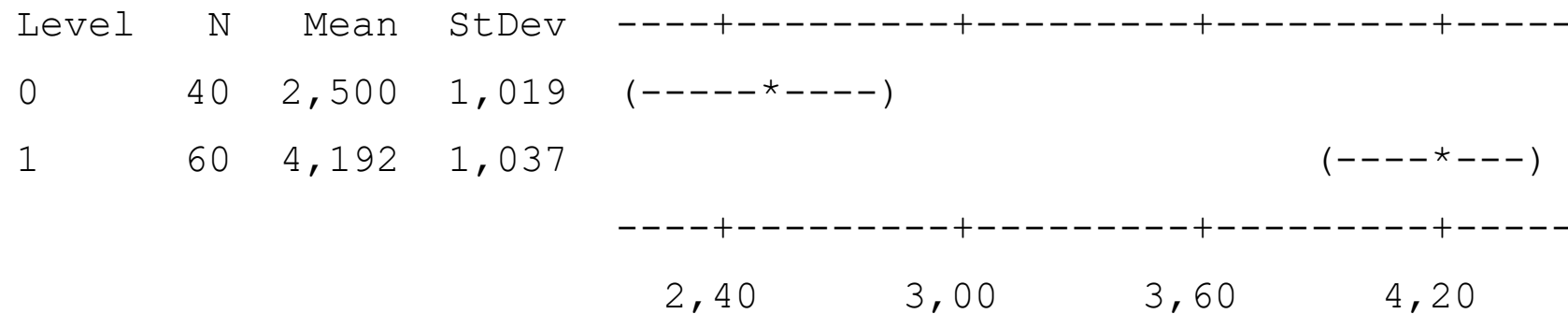


# One-way ANOVA: Delivery Speed versus Specification Buying

Source	DF	SS	MS	F	P
Specification Buying	1	68,68	68,68	64,72	0,000
Error	98	104,01	1,06		
Total	99	172,69			

S = 1,030    R-Sq = 39,77%    R-Sq(adj) = 39,16%

Individual 95% CIs For Mean Based on Pooled StDev



Pooled StDev = 1,030





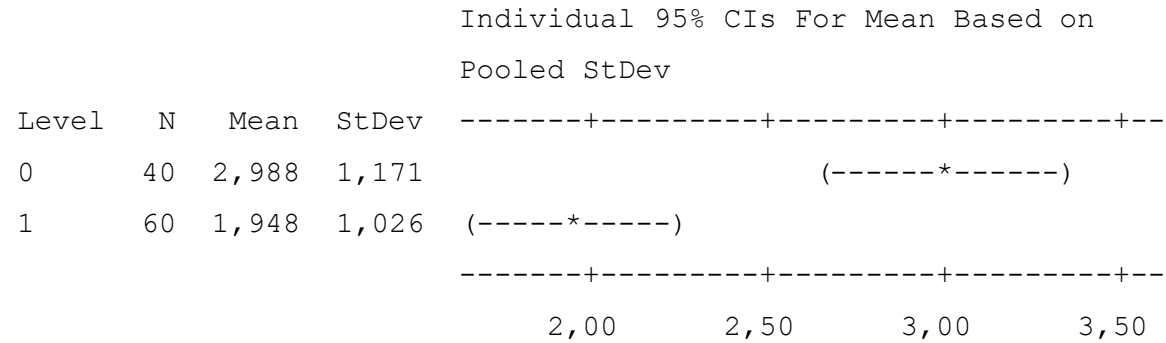
- Ao repetir o processo (ANOVA) para as demais variáveis (C3, C4, C5, C6, C7 e C8)



# One-way ANOVA: Price Level versus Specification Buying

Source	DF	SS	MS	F	P
Specification Buying	1	25,92	25,92	21,97	0,000
Error	98	115,61	1,18		
Total	99	141,53			

S = 1,086    R-Sq = 18,31%    R-Sq(adj) = 17,48%



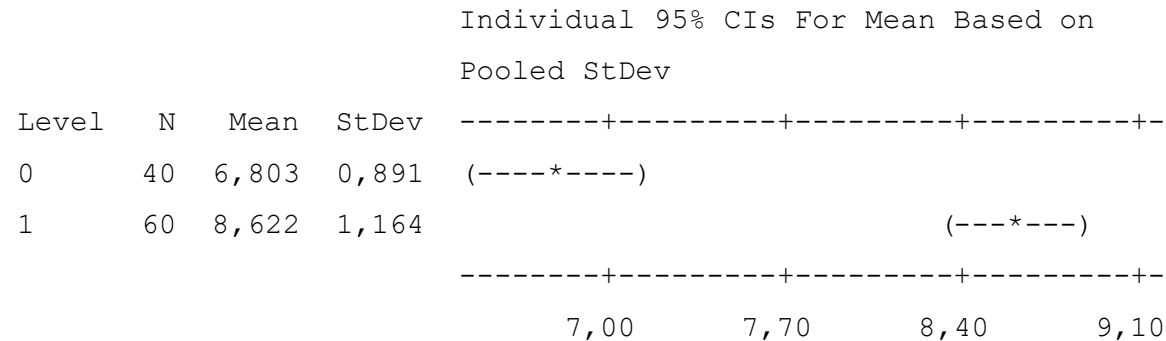
Pooled StDev = 1,086



# One-way ANOVA: Price Flexibility versus Specification Buying

Source	DF	SS	MS	F	P
Specification Buying	1	79,42	79,42	70,19	0,000
Error	98	110,89	1,13		
Total	99	190,32			

S = 1,064    R-Sq = 41,73%    R-Sq(adj) = 41,14%



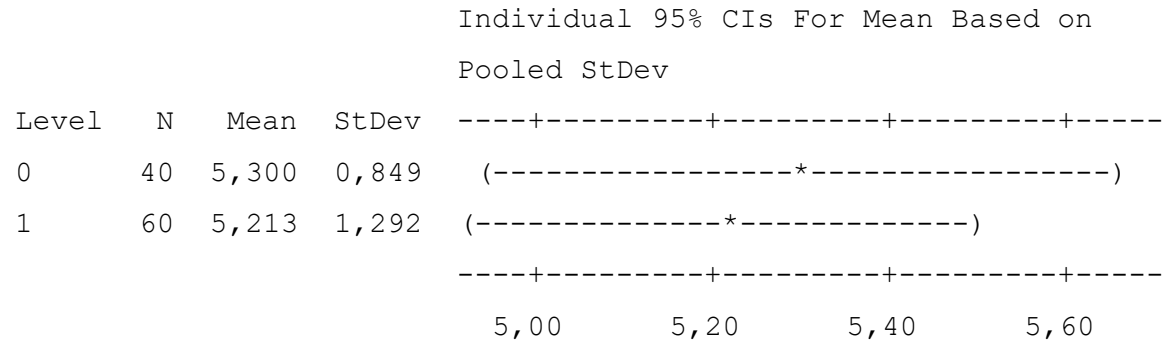
Pooled StDev = 1,064



# One-way ANOVA: Manufacturer Image versus Specification Buying

Source	DF	SS	MS	F	P
Specification Buying	1	0,18	0,18	0,14	0,709
Error	98	126,55	1,29		
Total	99	126,73			

S = 1,136    R-Sq = 0,14%    R-Sq(adj) = 0,00%



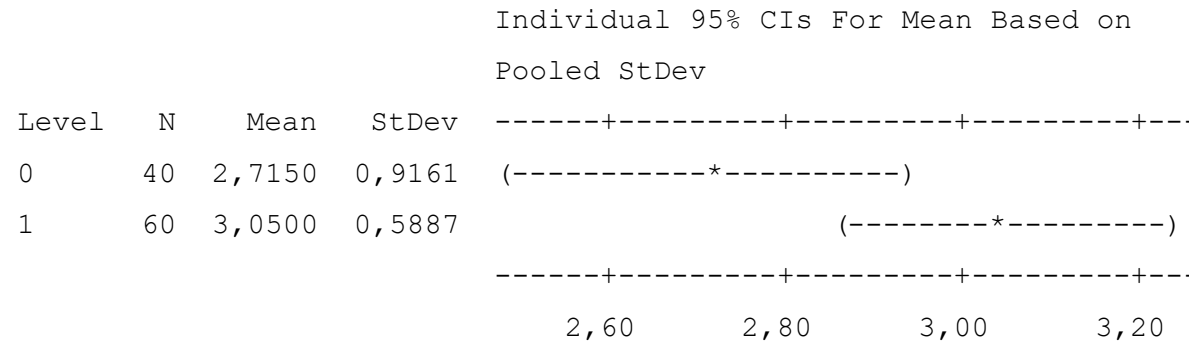
Pooled StDev = 1,136



# One-way ANOVA: Service versus Specification Buying

Source	DF	SS	MS	F	P
Specification Buying	1	2,693	2,693	4,96	0,028
Error	98	53,181	0,543		
Total	99	55,874			

S = 0,7367    R-Sq = 4,82%    R-Sq(adj) = 3,85%



Pooled StDev = 0,7367

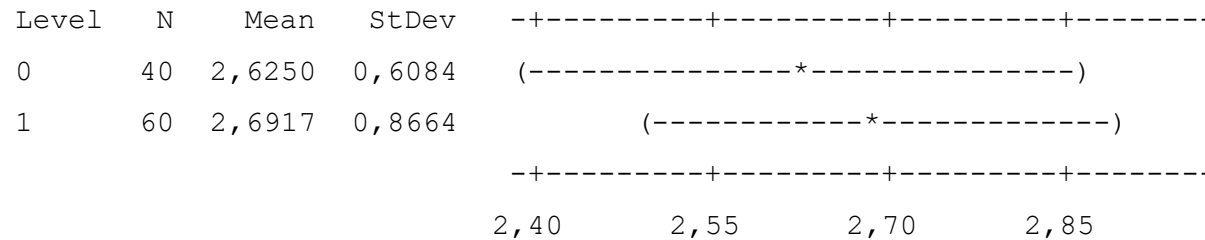


# One-way ANOVA: Salesforce Image versus Specification Buying

Source	DF	SS	MS	F	P
Specification Buying	1	0,107	0,107	0,18	0,674
Error	98	58,721	0,599		
Total	99	58,828			

S = 0,7741    R-Sq = 0,18%    R-Sq(adj) = 0,00%

Individual 95% CIs For Mean Based on Pooled StDev



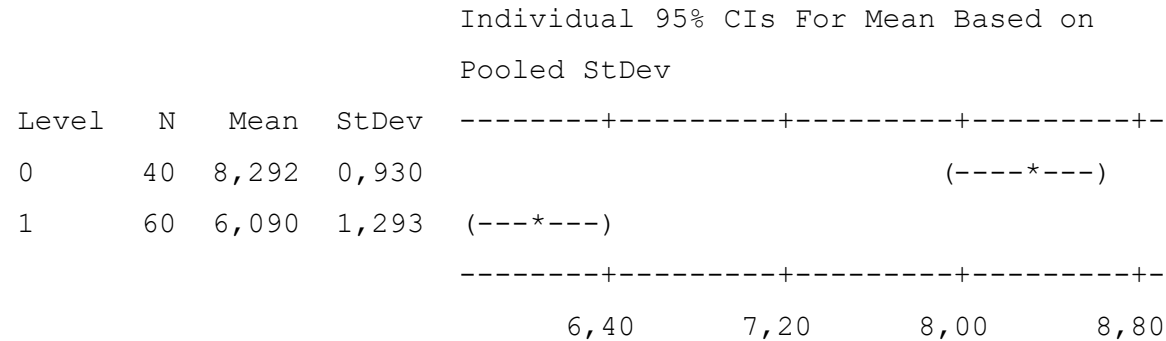
Pooled StDev = 0,7741



# One-way ANOVA: Product Quality versus Specification Buying

Source	DF	SS	MS	F	P
Specification Buying	1	116,42	116,42	86,20	0,000
Error	98	132,36	1,35		
Total	99	248,79			

S = 1,162    R-Sq = 46,80%    R-Sq(adj) = 46,25%



Pooled StDev = 1,162



# Síntese das ANOVAs

Variável	F	P
Delivery Speed	64,72	0,000
Price Level	21,97	0,000
Price Flexibility	70,19	0,000
Manufacturer Image	0,14	0,709
Service	4,96	0,028
Salesforce Image	0,18	0,674
Product Quality	86,20	0,000





- Basic Statistics
- Regression
- ANOVA
- DOE
- Control Charts
- Quality Tools
- Reliability/Survival
- Multivariate**
- Time Series
- Tables
- Nonparametrics
- EDA
- Power and Sample Size

- Principal Components...
- Factor Analysis...
- Item Analysis...
- Cluster Observations...
- Cluster Variables...
- Cluster K-Means...
- Discriminant Analysis...**
- Simple Correspondence Analysis...
- Multiple Correspondence Analysis...

Session

Predictors: Delivery Speed; Service  
 Group 0  
 Count 40

Summary of classification  
 Put into Group

	0	1
0	38	9
1	2	51
Total N	40	60
N correct	38	51
Proportion	0,950	0,850

N = 100      N Correct = 89      Proportion Correct = 0,890

Price Flexibility; Manufacturer Image;  
 Product Quality

FIRST FACTOR

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
	Id	Delivery Speed	Price Level	Price Flexibility	Manufacturer Image	Service	Salesforce Image	Product Quality	Firm Size	Usage Level	Satisfaction Level	Specification Buying
1	1	4	1	7	5	2	2	5	0	32	4	1
2	2	2	3	6	7	3	4	8	1	43	4	0
3	3	3	5	6	6	4	3	8	1	48	5	0
4	4	3	1	7	6	2	2	8	1	32	4	0
5	5	6	1	10	8	3	5	5	0	68	7	1



### Análise Discriminante

Grupos:

Preditores:

Função Discriminante  Usar validação cruzada  
 Linear  Quadrático

Armazenamento  
Função discriminante linear:

Ajustes  Ajustes de validação cruzada

Selecionar Ajuda Opções... OK Cancelar

### Análise Discriminante: Opções

Probabilidades a Priori:

Predizer identificadores de grupo para:

Exibição dos Resultados  
 Não exibir  
 Matriz de classificação  
 Acima mais fdl, distâncias e sumário de classificação errada  
 Acima mais média, desv. pad. e sumário de covariância  
 Acima mais sumário de classificação completo

Selecionar Ajuda OK Cancelar



Predictors: Delivery Speed; Price Level; Price Flexibility; Service; Product Quality

Group	0	1
Count	40	60

Summary of classification

	True Group	
Put into Group	0	1
0	36	10
1	4	50
Total N	40	60
N correct	36	50
Proportion	0,900	0,833

N=100      N Correct = 86      Proportion Correct = 0,860



# Linear Discriminant Function for Groups (Funções de Fisher)

Grupos	<u>0</u>	<u>1</u>
Constant	-54,029	-58,274
X1 - Delivery Speed	7,169	8,689
X2 - Price Level	8,407	8,862
X3 - Price Flexibility	6,810	8,243
X5 - Service	-9,967	-10,605
X7 - Product Quality	5,518	3,963



### Sumário de Observações Classificadas Incorretamente

Observação	Grupo		Distância		
	Verdadeiro	Pred Grupo	Grupo	Quadrática	Probabilidade
13**	0	1	0	4,317	0,341
			1	3,001	0,659
17**	1	0	0	6,884	0,782
			1	9,441	0,218
23**	1	0	0	7,669	0,528
			1	7,893	0,472
32**	1	0	0	6,139	0,588
			1	6,849	0,412
35**	0	1	0	5,825	0,457
			1	5,483	0,543
56**	1	0	0	1,401	0,798
			1	4,147	0,202
64**	1	0	0	6,741	0,827
			1	9,876	0,173
82**	1	0	0	8,502	0,809
			1	11,390	0,191
84**	1	0	0	2,672	0,830
			1	5,850	0,170
85**	0	1	0	5,051	0,289
			1	3,254	0,711
87**	0	1	0	5,172	0,349
			1	3,923	0,651
88**	1	0	0	2,429	0,862
			1	6,094	0,138
91**	1	0	0	2,149	0,806
			1	5,000	0,194
93**	1	0	0	9,318	0,856
			1	12,875	0,144



## Sumário de Observações Classificadas Incorretamente

Observação	Grupo		Grupo	Distância	
	Verdadeiro	Pred		Quadrática	Probabilidade
13**	0	1	0	4,317	0,341
			1	3,001	0,659
17**	1	0	0	6,884	0,782
			1	9,441	0,218
23**	1	0	0	7,669	0,528
			1	7,893	0,472
32**	1	0	0	6,139	0,588
			1	6,849	0,412
35**	0	1	0	5,825	0,457
			1	5,483	0,543
56**	1	0	0	1,401	0,798
			1	4,147	0,202



# Exemplo 2

## **Variável dependente com 3 categorias**

C15: situação compra

1: nova

2: primeira recompra

3: outras

## **Variáveis independentes**

C2: Delivery Speed

C3: Price Level

C4: Price Flexibility

C5: Manufacturer Image

C6: Service

C7: Salesforce Image

C8: Product Quality



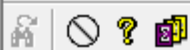
# Realizar Análise de Variância

O comportamento das variáveis que serão independentes na análise discriminante (C2, C3, C4, C5, C6, C7 e C8) muda nos grupos da variável Situação de Compra (C15)?





- Basic Statistics
- Regression
- ANOVA**
- DOE
- Control Charts
- Quality Tools
- Reliability/Survival
- Multivariate
- Time Series
- Tables
- Nonparametrics
- EDA
- Power and Sample Size



- One-Way...**
- One-Way (Unstacked)...
- Two-Way...
- Analysis of Means...
- Balanced ANOVA...
- General Linear Model...
- Fully Nested ANOVA...
- Balanced MANOVA...
- General MANOVA...
- Test for Equal Variances...
- Interval Plot...
- Main Effects Plot...
- Interactions Plot...



## Session

Level N Mean S

0	40	2,500	
1	60	4,192	

Pooled StDev = 1,00

### One-Way Analysis of Variance

Response: Delivery Speed

Factor: Buying Situation

Store residuals

Store fits

Confidence level: 95,0

Select Comparisons... Graphs...

Help OK Cancel

## Worksheet1 \*\*\*

↓	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
	Id	Delivery Speed	Price Level	Price Flexibility	Manufacturer Image	Service	Salesforce Image	Product Quality	Firm Size	Usage Level	Satisfaction Level	Specification Buying
1	1	4	1	7	5	2	2	5	0	32	4	1
2	2	2	3	6	7	3	4	8	1	43	4	0
3	3	3	5	6	6	4	3	8	1	48	5	0
4	4	3	1	7	6	2	2	8	1	32	4	0
5	5	6	1	10	8	3	5	5	0	58	7	0

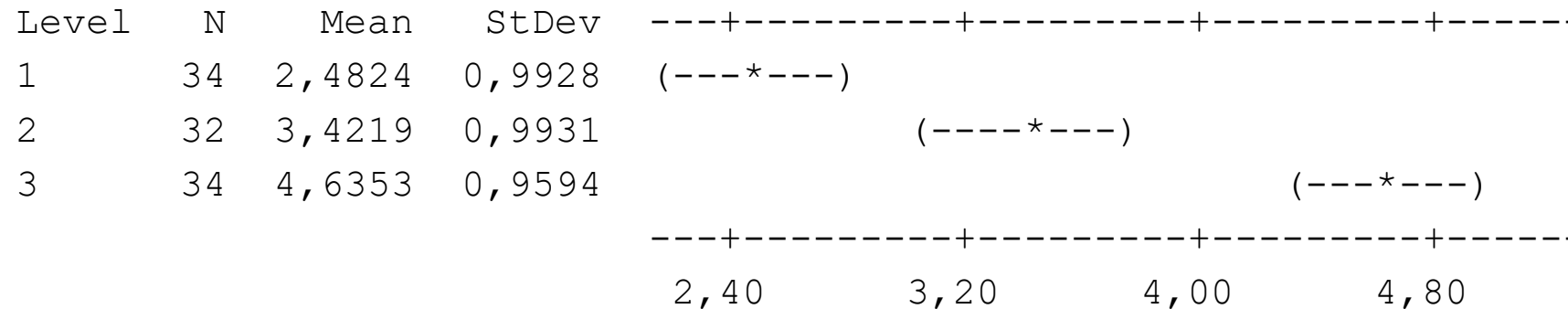


# One-way ANOVA: Delivery Speed versus Type of Buying Situation

Source	DF	SS	MS	F	P
Type of Buying Situation	2	79,206	39,603	41,09	0,000
Error	97	93,482	0,964		
Total	99	172,688			

S = 0,9817    R-Sq = 45,87%    R-Sq(adj) = 44,75%

Individual 95% CIs For Mean Based on Pooled StDev



Pooled StDev = 0,9817

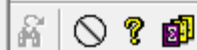


# Síntese das ANOVAs

Variável	F	P
Delivery Speed	41,09	0,000
Price Level	14,36	0,000
Price Flexibility	43,78	0,000
Manufacturer Image	2,44	0,092
Service	37,32	0,000
Salesforce Image	0,13	0,877
Product Quality	12,12	0,000



- Basic Statistics
- Regression
- ANOVA
- DOE
- Control Charts
- Quality Tools
- Reliability/Survival
- Multivariate
- Time Series
- Tables
- Nonparametrics
- EDA
- Power and Sample Size



## Session

Predictors: Delivery Speed; Service

Group 0  
Count 40

Summary of classification

Put into Group	0	1
0	38	9
1	2	51
Total N	40	60
N correct	38	51
Proportion	0,950	0,850

N = 100

N Correct = 89

Price Flexibility; Manufacturer Image; Product Quality

Principal Components...

Factor Analysis...

Item Analysis...

Cluster Observations...

Cluster Variables...

Cluster K-Means...

Discriminant Analysis...

Simple Correspondence Analysis...

Multiple Correspondence Analysis...

## FIRST FACTOR

↓	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
	Id	Delivery Speed	Price Level	Price Flexibility	Manufacturer Image	Service	Salesforce Image	Product Quality	Firm Size	Usage Level	Satisfaction Level	Specification Buying
1	1	4	1	7	5	2	2	5	0	32	4	1
2	2	2	3	6	7	3	4	8	1	43	4	0
3	3	3	5	6	6	4	3	8	1	48	5	0
4	4	3	1	7	6	2	2	8	1	32	4	0
5	5	6	1	10	8	3	5	5	0	58	7	1



**Discriminant Analysis**

Groups:

Predictors:

Discriminant Function  Use cross validation  
 Linear  Quadratic

Storage  
Linear discriminant function:

Fits  Fits from cross validation

Select Help Options... OK Cancel

**Discriminant Analysis - Options**

Prior probabilities:

Predict group membership for:

Display of Results  
 Do not display  
 Classification matrix  
 Above plus |df, distances, and misclassification summary  
 Above plus mean, std. dev., and covariance summary  
 Above plus complete classification summary

Select Help OK Cancel



# Discriminant Analysis: Type of Buyi versus Delivery Spe; Price Level; ...

Linear Method for Response: Type of Buying Situation

Predictors: Delivery Speed; Price Level; Price Flexibility; Service; Product Quality

Group	1	2	3
Count	34	32	34

Summary of classification

Put into Group	True Group		
	1	2	3
1	27	4	0
2	5	21	2
3	2	7	32
Total N	34	32	34
N correct	27	21	32
Proportion	0,794	0,656	0,941

N = 100    N Correct = 80    Proportion Correct = 0,800



# Linear Discriminant Function for Groups (Funções de Fisher)

Grupos	1	2	3
Constant	-67,147	-81,524	-98,631
Delivery Speed	6,181	6,736	5,432
Price Level	6,538	7,673	5,414
Price Flexibility	10,416	11,354	13,100
Service	-1,105	1,235	5,839
Product Quality	4,386	3,921	3,709



# Exercício – Análise Discriminante

## Cereais Matinais

### Variáveis

- Brand
- Manufacturer : G, K e Q
- Calories
- Protein
- Fat
- Sodium
- Fiber
- Carbohydrates
- Sugar
- Potassium

Construir um modelo discriminante

Variável Dependente: Manufacturer (G, K e Q)

Variáveis Independentes

- Calories
- Protein
- Fat
- Sodium
- Fiber
- Carbohydrates
- Sugar
- Potassium