AULA 28
Correlação e Análise Fatorial

Ernesto F. L. Amaral

21 de junho de 2011
Avaliação de Políticas Públicas (DCP 046)

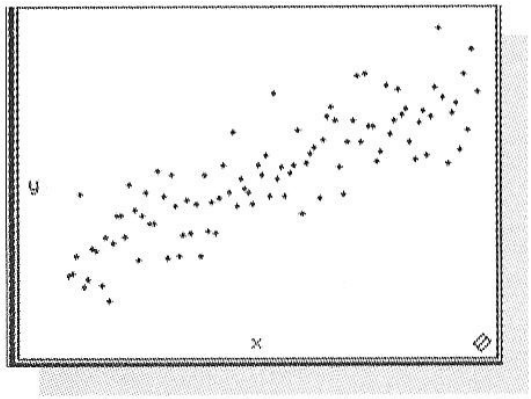
CORRELAÇÃO

CONCEITOS BÁSICOS

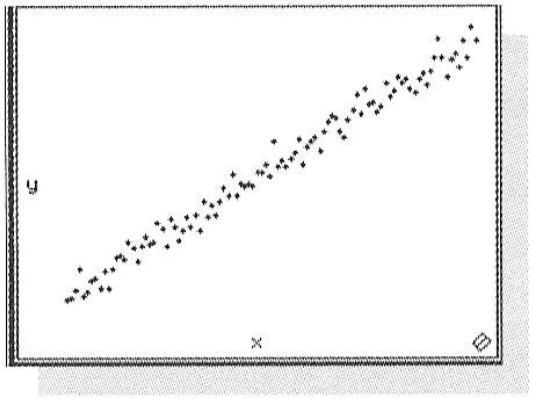
– Existe uma correlação entre duas variáveis quando uma delas está relacionada com a outra de alguma maneira.

– Antes de tudo é importante explorar os dados:
  – Diagrama de dispersão entre duas variáveis.
  – Há tendência?
  – Crescente ou decrescente?
  – *Outliers*?

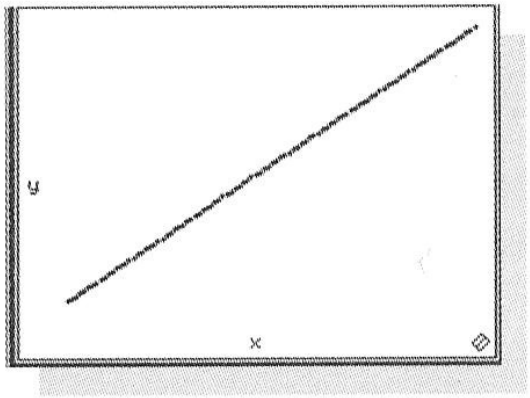
DIAGRAMAS DE DISPERSÃO
(correlação linear)



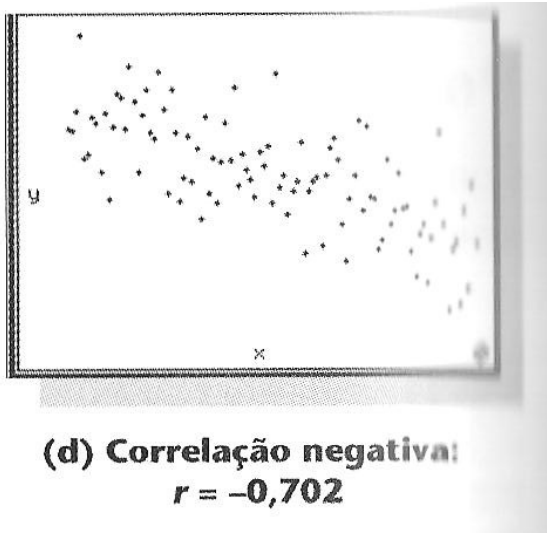
(a) Correlação positiva:
$r = 0{,}851$



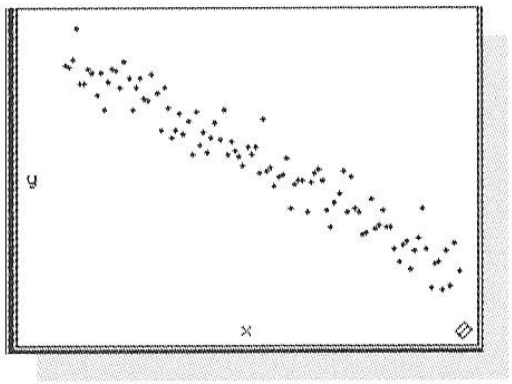
(b) Correlação positiva:
$r = 0{,}991$



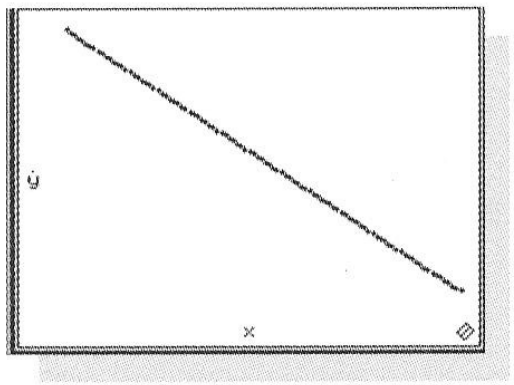
(c) Correlação positiva perfeita:
$r = 1$



(d) Correlação negativa:
$r = -0{,}702$

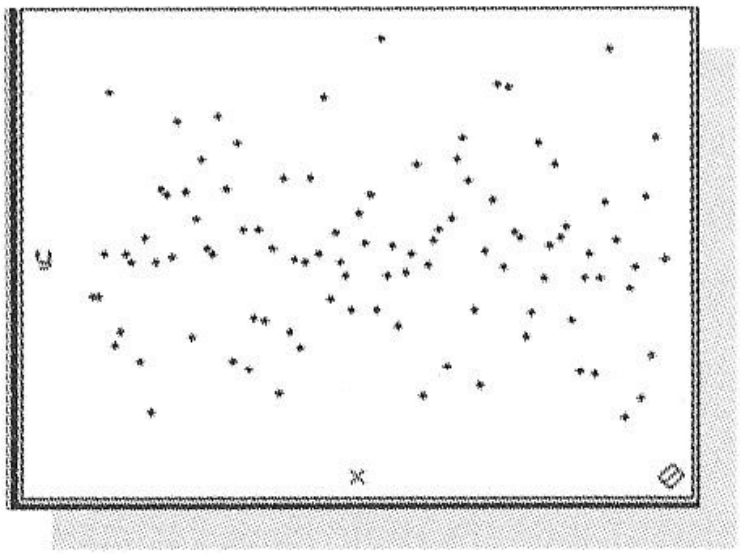


(e) Correlação negativa:
$r = -0{,}965$

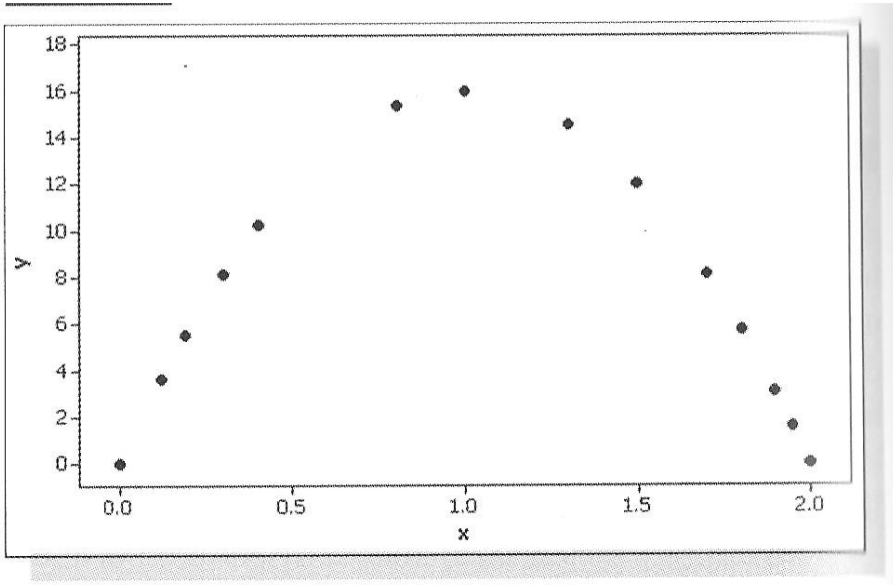


(f) Correlação negativa perfeita:
$r = -1$

DIAGRAMAS DE DISPERSÃO
(não há correlação linear)



(g) Nenhuma correlação: $r = 0$



(h) Relação não-linear: $r = -0{,}087$

CORRELAÇÃO

– O coeficiente de correlação linear (*r*):

– Medida numérica da força da relação entre duas variáveis que representam dados quantitativos.

– Mede intensidade da relação linear entre os valores quantitativos emparelhados *x* e *y* em uma amostra.

– É chamado de coeficiente de correlação do produto de momentos de Pearson.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2}\sqrt{n(\sum y^2) - (\sum y)^2}}$$

OBSERVAÇÕES IMPORTANTES

– Usando dados amostrais emparelhados (dados bivariados), estimamos valor de *r* para concluir se há ou não relação entre duas variáveis.

– Serão tratadas relações lineares, em que pontos no gráfico (x, y) se aproximam do padrão de uma reta.

– É importante entender os conceitos e não os cálculos aritméticos.

– *r* é calculado com dados amostrais. Se tivéssemos todos pares de valores populacionais *x* e *y*, teríamos um parâmetro populacional (ρ).

REQUISITOS

– Os seguintes requisitos devem ser satisfeitos ao se testarem hipóteses ou ao se fazerem outras inferências sobre *r* :

- – Amostra de dados emparelhados (*x*, *y*) é uma **amostra aleatória** de dados quantitativos independentes.
  - – Não pode ter sido utilizado, por exemplo, amostra de resposta voluntária.
- – Exame visual do diagrama de dispersão deve confirmar que pontos se aproximam do **padrão de uma reta**.
- – **Valores extremos** (*outliers*) devem ser removidos se forem erros.
  - – Efeitos de outros *outliers* devem ser considerados com estimação de *r* com e sem estes *outliers*.

VALORES CRÍTICOS DO COEFICIENTE DE CORRELAÇÃO DE PEARSON (*r*)

n	$\alpha = 0{,}05$	$\alpha = 0{,}01$
4	0,950	0,999
5	0,878	0,959
6	0,811	0,917
7	0,754	0,875
8	0,707	0,834
9	0,666	0,798
10	0,632	0,765
11	0,602	0,735
12	0,576	0,708
13	0,553	0,684
14	0,532	0,661
15	0,514	0,641
16	0,497	0,623
17	0,482	0,606
18	0,468	0,590
19	0,456	0,575
20	0,444	0,561
25	0,396	0,505
30	0,361	0,463
35	0,335	0,430
40	0,312	0,402
45	0,294	0,378
50	0,279	0,361
60	0,254	0,330
70	0,236	0,305
80	0,220	0,286
90	0,207	0,269
100	0,196	0,256

NOTA: Para testar H_0: $\rho = 0$ *versus* H_1: $\rho \neq 0$, rejeite H_0 se o valor absoluto de *r* for maior que o valor crítico na tabela.

– **Arredonde** o coeficiente de correlação linear *r* para três casas decimais, permitindo comparação com esta tabela.

– Interpretação: com 4 pares de dados e **nenhuma correlação** linear entre *x* e *y*, há chance de 5% de que valor absoluto de *r* exceda 0,950.

INTERPRETANDO *r*

– O valor de *r* deve sempre estar entre –1 e +1.

– Se *r* estiver muito próximo de 0, concluímos que não há correlação linear significativa entre *x* e *y*.

– Se *r* estiver próximo de –1 ou +1, concluímos que há uma relação linear significativa entre *x* e *y*.

– Mais objetivamente:

  – Usando a tabela anterior, se valor absoluto de *r* excede o valor da tabela, há correlação linear.

  – Usando programa de computador, se valor *P* é menor do que nível de significância, há correlação linear.

PROPRIEDADES DE *r*

– Valor de *r* está entre: $-1 \leq r \leq +1$

– Valor de *r* não muda se todos valores de qualquer das variáveis forem convertidos para uma escala diferente.

– Valor de *r* não é afetado pela inversão de *x* ou *y*. Ou seja, mudar os valores de *x* pelos valores de *y* e vice-versa não modificará *r*.

– *r* mede intensidade de relação linear, não sendo planejado para medir intensidade de relação que não seja linear.

– O valor de r^2 é a proporção da variação em *y* que é explicada pela relação linear entre *x* e *y*.

ERROS DE INTERPRETAÇÃO

– Erro comum é concluir que correlação implica **causalidade**:
  – A causa pode ser uma variável oculta.
  – Uma variável oculta é uma variável que afeta as variáveis em estudo, mas que não está incluída no banco.

– Erro surge de dados que se baseiam em **médias**:
  – Médias suprimem variação individual e podem aumentar coeficiente de correlação.

– Erro decorrente da propriedade de **linearidade**:
  – Pode existir relação entre *x* e *y* mesmo quando não haja correlação linear (relação quadrática, por exemplo).

TESTE DE HIPÓTESE FORMAL PARA CORRELAÇÃO

– É possível realizar um teste de hipótese formal para determinar se há ou não relação linear significativa entre duas variáveis.

– Critério de decisão é rejeitar a hipótese nula (ρ=0) se o valor absoluto da estatística de teste exceder os valores críticos.

– A rejeição de (ρ=0) significa que há evidência suficiente para apoiar a afirmativa de uma correlação linear entre as duas variáveis.

– Se o valor absoluto da estatística de teste não exceder os valores críticos (ou seja, o valor *P* for grande), deixamos de rejeitar ρ=0.

H_0: ρ=0 (não há correlação linear)

H_1: $\rho \neq 0$ (há correlação linear)

MÉTODO 1: ESTATÍSTICA DE TESTE É *t*

– Estatística de teste representa o valor do desvio padrão amostral dos valores de *r* :

$$t = \frac{r}{\sqrt{\frac{1 - r^2}{n - 2}}}$$

– Valores críticos e valor *P*: use tabela A-3 com *n*–2 graus de liberdade.

– Conclusão:

  – Se |*t*| > valor crítico da Tabela A-3, rejeite H_0 e conclua que há correlação linear.

  – Se |*t*| ≤ valor crítico da Tabela A-3, deixe de rejeitar H_0 e conclua que não há evidência suficiente para concluir que haja correlação linear.

MÉTODO 2: ESTATÍSTICA DE TESTE É *r*

– Estatística de teste: *r*

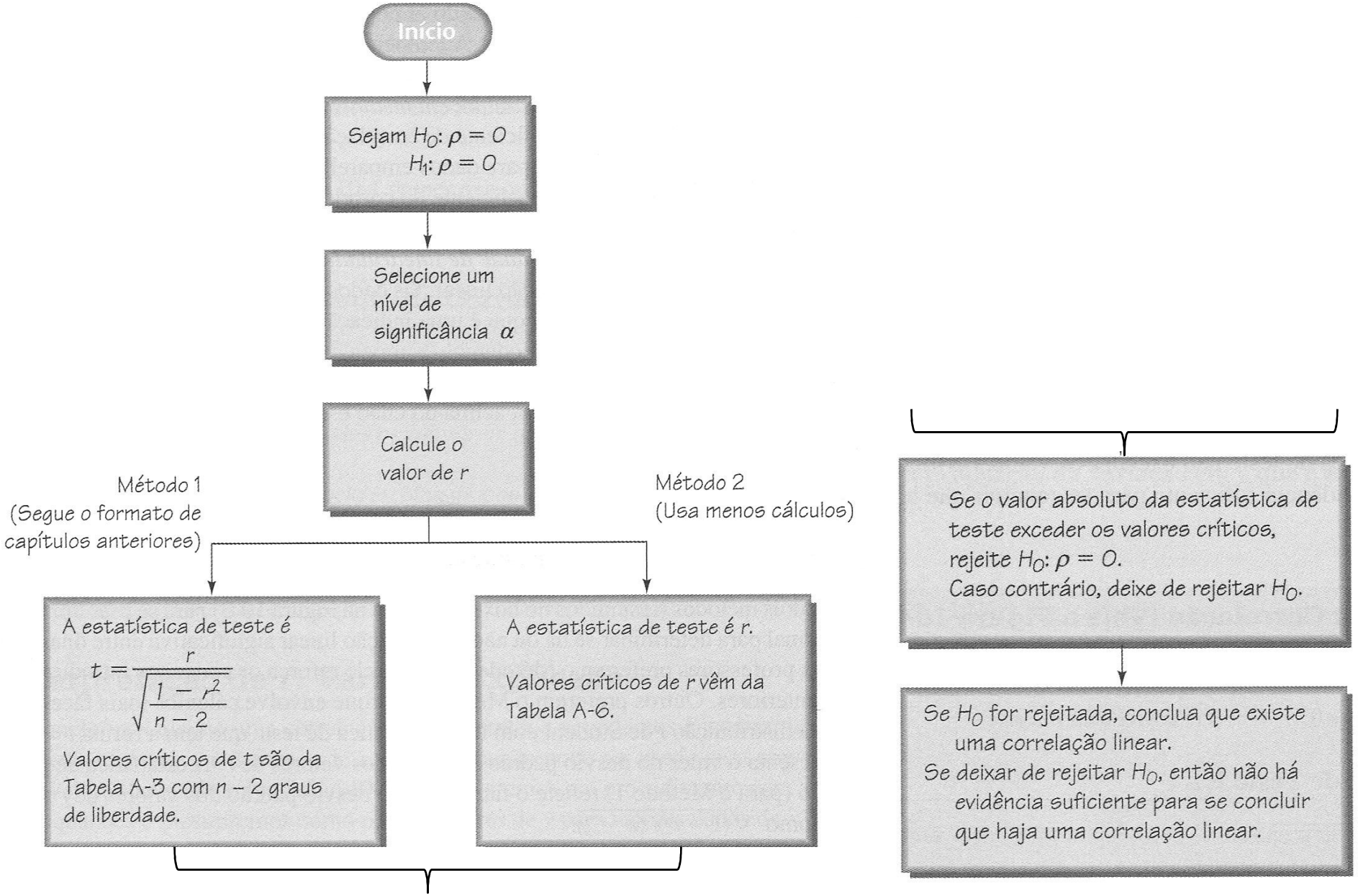
– Valores críticos: consulte Tabela A-6.

– Conclusão:

  – Se $|r|$ > valor crítico da Tabela A-6, rejeite H_0 e conclua que há correlação linear.

  – Se $|r|$ ≤ valor crítico da Tabela A-6, deixe de rejeitar H_0 e conclua que não há evidência suficiente para concluir que haja correlação linear.

TESTE DE HIPÓTESE PARA CORRELAÇÃO LINEAR



TESTES UNILATERAIS

– Os testes unilaterais podem ocorrer com uma afirmativa de uma correlação linear positiva ou uma afirmativa de uma correlação linear negativa.

– Afirmativa de correlação negativa (teste unilateral esquerdo):

$$H_0: \rho = 0$$

$$H_1: \rho < 0$$

– Afirmativa de correlação positiva (teste unilateral direito):

$$H_0: \rho = 0$$

$$H_1: \rho > 0$$

– Para isto, simplesmente utilize α=0,025 (ao invés de α=0,05) e α=0,005 (ao invés de α=0,01).

FUNDAMENTOS

– Essas fórmulas são diferentes versões da mesma expressão:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2}\sqrt{n(\sum y^2) - (\sum y)^2}}$$

$$r = \frac{\sum\left[\frac{(x - \bar{x})}{s_x}\frac{(y - \bar{y})}{s_y}\right]}{n - 1}$$

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{(n - 1)s_x s_y} \qquad r = \frac{s_{xy}}{\sqrt{s_{xx}}\sqrt{s_{yy}}}$$

FUNDAMENTOS

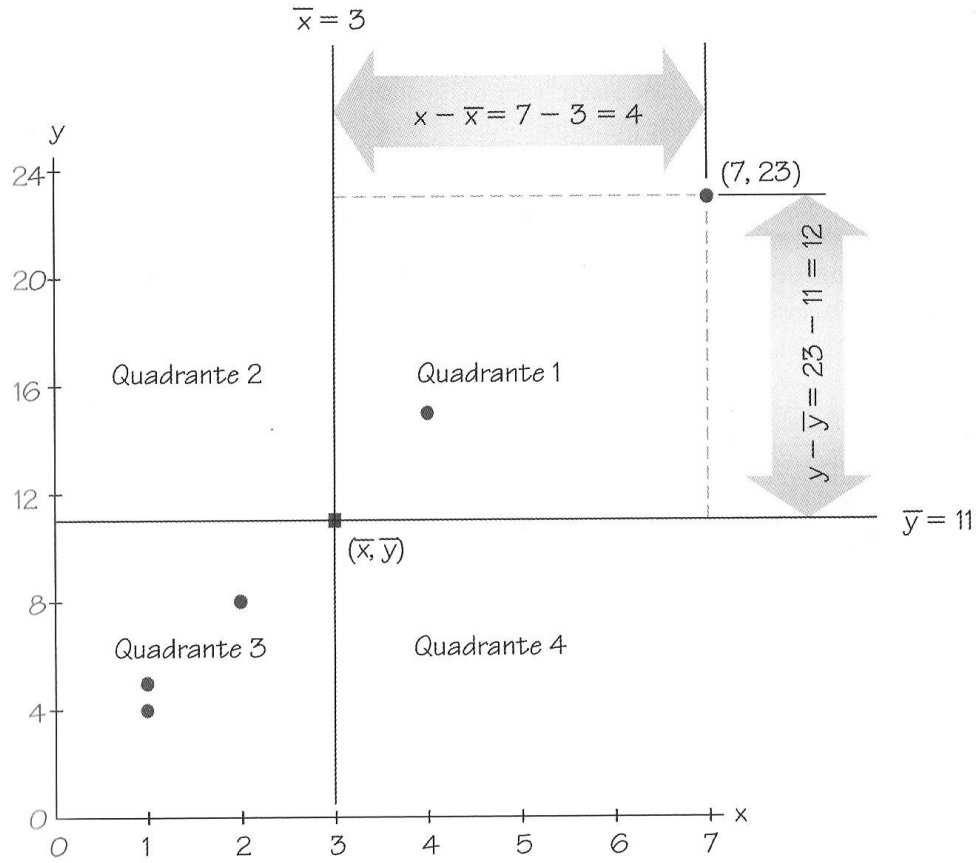
– Dada uma coleção de dados em pares (*x*,*y*), o ponto (*x*-barra, *y*-barra) é chamado de **centróide**.

– A estatística do produto dos momentos de Pearson (*r*) se baseia na soma dos produtos dos momentos:

$$\sum (x - \bar{x})(y - \bar{y})$$

– Se pontos são reta ascendente, valores do produto estarão nos 1º e 3º quadrantes (soma positiva).

– Se é descendente, os pontos estarão nos 2º e 4º quadrantes (soma negativa).



OU SEJA...

– Podemos usar esta expressão para medir como pontos estão organizados:

$$\sum (x - \bar{x})(y - \bar{y})$$

– Grande soma positiva sugere pontos predominantemente no primeiro e terceiro quadrantes (correlação linear positiva).

– Grande soma negativa sugere pontos predominantemente no segundo e quarto quadrantes (correlação linear negativa).

– Soma próxima de zero sugere pontos espalhados entre os quatro quadrantes (não há correlação linear).

PORÉM...

– Esta soma depende da magnitude dos números usados:

$$\sum (x - \bar{x})(y - \bar{y})$$

– Para tornar *r* independente da escala utilizada, usamos a seguinte padronização:

  – Sendo s_x o desvio padrão dos valores amostrais *x*...

  – Sendo s_y o desvio padrão dos valores amostrais *y*...

  – Padronizamos cada desvio pela sua divisão por s_x...

$$\sum \left[\frac{(x - \bar{x})}{s_x} \frac{(y - \bar{y})}{s_y} \right]$$

– Usamos o divisor $n - 1$ para obter uma espécie de média:

$$r = \frac{\sum \left[\frac{(x - \bar{x})}{s_x} \frac{(y - \bar{y})}{s_y} \right]}{n - 1}$$

COMANDOS NO STATA

– Podemos usar os comandos “correlate” ou “pwcorr”, em que ambos mostram a matriz de correlações entre as variáveis.

– O comando “corr” usa “listwise deletion”, em que toda matriz é calculada somente para casos que não possuem nenhum valor em branco (*missing*) em nenhuma variável na lista:

corr x y z

– O comando “pwcorr” usa “pairwise deletion”, em que cada correlação é computada para casos que não possuem nenhum valor em branco para cada par de variáveis:

pwcorr x y z, sig

– Uso do “pwcorr” para obter o mesmo que “corr”:

pwcorr x y z if !missing(x, y, z), sig

ANÁLISE FATORIAL

ANÁLISE FATORIAL

– Os **fatores** ou **construtos** são variáveis hipotéticas, combinações lineares das variáveis observadas, que explicam partes da variabilidade dos dados.

– **Análise fatorial** é usada principalmente com o objetivo de simplificar os dados:

1) Pegar um pequeno número de variáveis (preferencialmente não-correlacionadas) de um grande número de variáveis (em que maioria é correlacionada com a outra).

2) Criar índices com variáveis que medem dimensões conceituais similares.

TIPOS DE ANÁLISE FATORIAL

– **Exploratória:** quando não temos uma idéia pré-definida da estrutura e de quantas dimensões estão presentes em um conjunto de variáveis.

```
factor var1 var2 var3 ... varn
```

– **Confirmatória:** quando queremos testar hipóteses específicas sobre a estrutura de um número de dimensões subjacentes a um conjunto de variáveis. Por exemplo, pensamos que nossos dados possuem duas dimensões e queremos verificar isto.

```
factor var1 var2 var3 ... varn, factors(#)
```

– **Ajuda no Stata:**

```
help factor
```

MATRIZES

– A **matriz de correlação** é uma matriz quadrada cujos elementos são as correlações entre as variáveis analisadas.

– Na diagonal principal todos os elementos são iguais a 1 (um), visto que cada variável é totalmente correlacionada com ela mesma.

– A **matriz de covariância** é uma matriz quadrada cujos elementos fora da diagonal principal são as covariâncias entre as variáveis e na diagonal principal são as variâncias de cada variável.

MATRIZ DE CORRELAÇÃO (“corr” NO STATA)

```
. corr terpro aguarg escoarg lixod eletrica
(obs=134)

             |   terpro   aguarg  escoarg    lixod eletrica
-------------+------------------------------------------------
      terpro |   1.0000
      aguarg |   0.0021   1.0000
     escoarg |  -0.0192   0.2343   1.0000
       lixod |   0.1246   0.4393   0.5296   1.0000
    eletrica |   0.1214   0.2126   0.1253   0.1435   1.0000
```

COMPONENTES DA ANÁLISE FATORIAL

```
. factor terpro aguarg escoarg lixod eletrica, pcf
(obs=134)

Factor analysis/correlation                       Number of obs    =      134
    Method: principal-component factors           Retained factors =        2
    Rotation: (unrotated)                         Number of params =        9
    ------------------------------------------------------------------------------
         Factor  |   Eigenvalue   Difference        Proportion   Cumulative
    -------------+----------------------------------------------------------------
        Factor1  |      1.90885      0.84141            0.3818       0.3818
        Factor2  |      1.06744      0.17154            0.2135       0.5953
        Factor3  |      0.89590      0.15994            0.1792       0.7744
        Factor4  |      0.73596      0.34410            0.1472       0.9216
        Factor5  |      0.39186            .            0.0784       1.0000
    ------------------------------------------------------------------------------
    LR test: independent vs. saturated:  chi2(10) =   84.37 Prob>chi2 = 0.0000

Factor loadings (pattern matrix) and unique variances
    ------------------------------------------------
        Variable |  Factor1   Factor2 |   Uniqueness
    -------------+--------------------+--------------
          terpro |   0.1573    0.8300 |      0.2864
          aguarg |   0.6914   -0.0594 |      0.5185
         escoarg |   0.7233   -0.3011 |      0.3862
           lixod |   0.8428   -0.1055 |      0.2786
        eletrica |   0.4155    0.5228 |      0.5541
    ------------------------------------------------
```

AUTOVALORES & DIFERENÇA

– Os **autovalores (*eigenvalues*)** são valores obtidos a partir das matrizes de covariância ou de correlação, cujo objetivo é obter um conjunto de vetores independentes, não correlacionados, que expliquem o máximo da variabilidade dos dados.

– Indicam o total da variância causado por cada fator.

– A soma de todos autovalores é igual ao número de variáveis. Quando há valores negativos, a soma dos autovalores é igual ao número total de variáveis com valores positivos.

– O **critério de *Kaiser*** sugere utilizar os fatores com autovalores iguais ou superiores a uma unidade.

– **Diferença (*difference*)** é a subtração entre um autovalor e o próximo autovalor.

PROPORÇÃO & CUMULATIVA

– **Proporção (*proportion*)** indica o peso relativo de cada fator na variância total (variabilidade) dos dados.

– **Proporção cumulativa (*cumulative*)** indica o total da variância explicado por n+(n-1) fatores.

CARGAS FATORIAIS

– **Cargas fatoriais (*factors*)** são as correlações entre as variáveis originais e os fatores.

– Esse é um dos pontos principais da análise fatorial, quanto maior a carga fatorial maior será a correlação com determinado fator.

– Um valor negativo indica um impacto inverso no fator.

– A quantidade de cargas fatoriais é automaticamente calculada pelo Stata com base nos autovalores (iguais ou superiores a uma unidade - critério de *Kaiser*).

– Por sua vez, as **cargas fatoriais relevantes** são aquelas com valores **maiores que 0,5**.

COMUNALIDADE & ESPECIFICIDADE

– As **comunalidades (*communalities*)** são quantidades das variâncias (correlações) de cada variável explicada pelos fatores. Quanto maior a comunalidade, maior será o poder de explicação daquela variável pelo fator.

Desejamos comunalidades superiores a 0,5.

– A **especificidade** (***uniqueness***) ou erro é a parcela da variância (correlação) dos dados que não pode ser explicada pelo fator. É a proporção única da variável não compartilhada com as outras variáveis. É igual a 1 menos a comunalidade. Quanto maior a especificidade, menor é a relevância da variável no modelo fatorial.

Desejamos especificidades inferiores a 0,5.

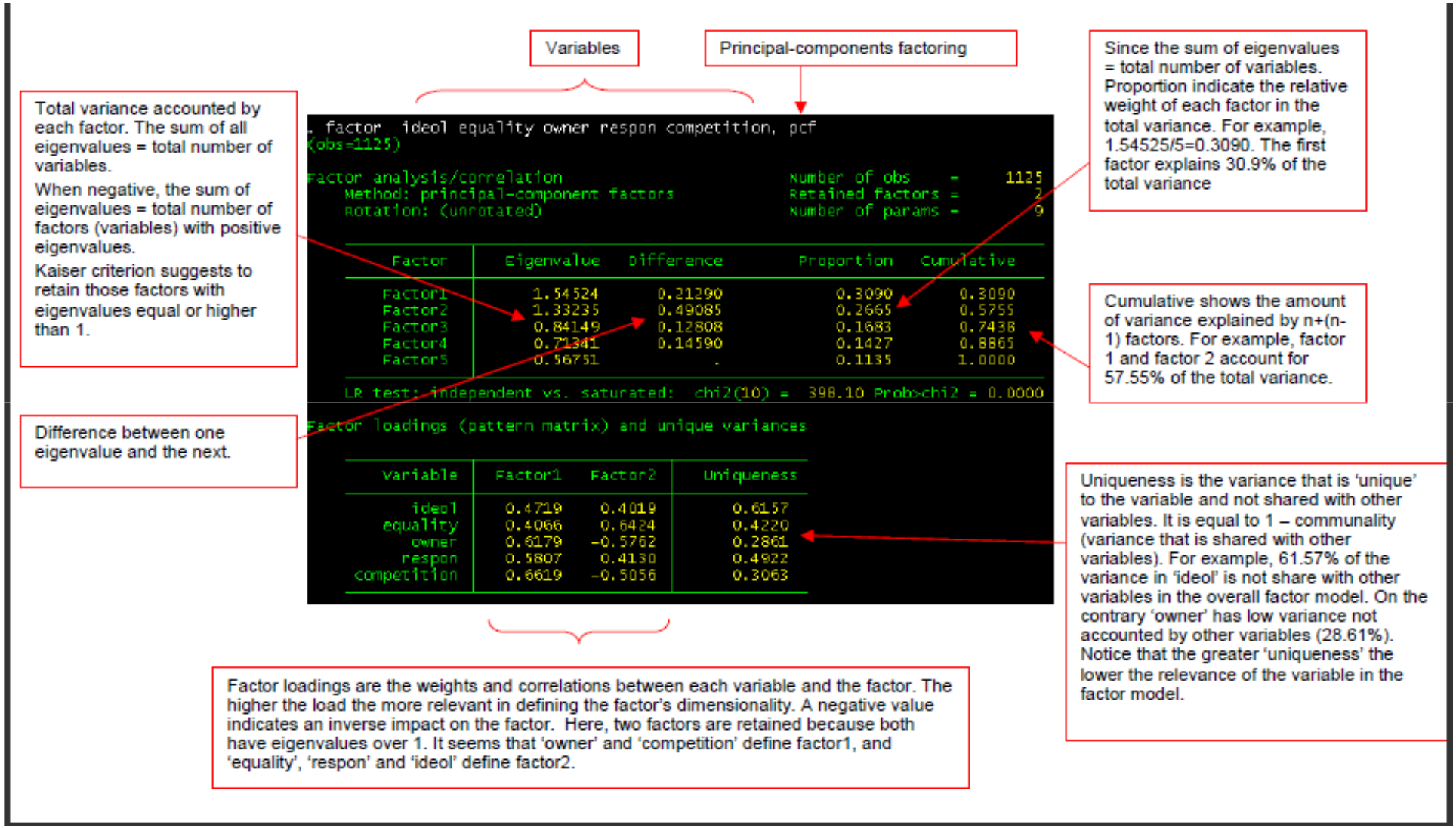
ANÁLISE DE COMPONENTES PRINCIPAIS

– Na **análise de componentes principais - ACP (*principal component factors - PCF*)**, quase todas variáveis estão altamente correlacionadas ao primeiro fator.

– Ou seja, é definido que a primeira componente (fator) explique a maior parte da variabilidade dos dados e por conseqüência as variáveis estarão mais correlacionadas a ela.

```
factor var1 var2 var3 ... varn, pcf
```

factor ideol equality owner respon competition, pcf



. factor ideol equality owner respon competition, pcf
(obs=1125)

Factor analysis/correlation — Number of obs = 1125
Method: principal-component factors — Retained factors = 2
Rotation: (unrotated) — Number of params = 9

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor1	1.54524	0.21290	0.3090	0.3090
Factor2	1.33235	0.49085	0.2665	0.5755
Factor3	0.84149	0.12808	0.1683	0.7438
Factor4	0.71341	0.14590	0.1427	0.8865
Factor5	0.56751	.	0.1135	1.0000

LR test: independent vs. saturated: chi2(10) = 398.10 Prob>chi2 = 0.0000

Factor loadings (pattern matrix) and unique variances

Variable	Factor1	Factor2	Uniqueness
ideol	0.4719	0.4019	0.6157
equality	0.4066	0.6424	0.4220
owner	0.6179	-0.5762	0.2861
respon	0.5807	0.4130	0.4922
competition	0.6619	-0.5056	0.3063

Fonte: Torres-Reyna, Oscar. s.d. *Getting Started in Factor Analysis (using Stata).* (http://dss.princeton.edu/training/)

ROTAÇÃO FATORIAL

– A **ACP** é um método estatístico multivariado que permite transformar um conjunto de variáveis iniciais correlacionadas entre si, num outro conjunto de variáveis não-correlacionadas (ortogonais), as chamadas componentes principais, que resultam de combinações lineares do conjunto inicial.

– Uma **rotação fatorial (*rotation*)** é o processo de manipulação ou de ajuste dos eixos fatoriais para conseguir uma solução fatorial mais simples e pragmaticamente mais significativa, cujos fatores sejam mais facilmente interpretáveis.

– A nova matriz padrão apresenta de forma mais clara a relevância de cada variável em cada fator.

`rotate`

rotate

By default the rotation is varimax which produces orthogonal factors. This means that factors are not correlated to each other. This setting is recommended when you want to identify variables to create indexes or new variables without inter-correlated components

Same description as in the previous slide with new composition between the two factors. Still both factors explain 57.55% of the total variance observed.

The pattern matrix here offers a clearer picture of the relevance of each variable in the factor. Factor1 is mostly defined by 'owner' and 'competition' and factor2 by 'equality', 'respon' and 'ideol' .

This is a correlation matrix between factor1 and factor2.

```
. rotate

Factor analysis/correlation                       Number of obs    =      1125
    Method: principal-component factors           Retained factors =         2
    Rotation: orthogonal varimax (Kaiser off)     Number of params =         9

    Factor  |   Variance   Difference    Proportion   Cumulative
------------+------------------------------------------------------------
    Factor1 |    1.45169      0.02579       0.2903       0.2903
    Factor2 |    1.42590            .       0.2852       0.5755

LR test: independent vs. saturated:  chi2(10) =  398.10 Prob>chi2 = 0.0000

Rotated factor loadings (pattern matrix) and unique variances

    Variable |   Factor1    Factor2   Uniqueness
-------------+--------------------------------
       ideol |     0.0869     0.6138       0.6157
    equality |    -0.1214     0.7505       0.4220
       owner |     0.8446    -0.0218       0.2861
      respon |     0.1610     0.6941       0.4922
 competition |     0.8307     0.0603       0.3063

Factor rotation matrix

             |    Factor1    Factor2
-------------+--------------------
    Factor1  |     0.7487     0.6629
    Factor2  |    -0.6629     0.7487
```

NOTE: If you want the factors to be correlated (oblique rotation) you need to use the option `promax` after `rotate`:

```
rotate, promax
```

Type `help rotate` for details.

Fonte: Torres-Reyna, Oscar. s.d. *Getting Started in Factor Analysis (using Stata).* (http://dss.princeton.edu/training/)

TESTE KAISER-MEYER-OLKLIN

– O teste de Kaiser-Meyer-Olklin (KMO) varia entre 0 e 1. Quanto mais perto de 1, melhor.

– Friel (2009) sugere a seguinte escala para interpretar o valor da estatística KMO:

* Entre 0,90 e 1: excelente.
* Entre 0,80 e 0,89: bom.
* Entre 0,70 e 0,79: mediano.
* Entre 0,60 e 0,69: medíocre.
* Entre 0,50 e 0,59: ruim.
* Entre 0 e 0,49: inadequado.

– Pallant (2007) sugere 0,60 como um limite razoável.

– Hair et al. (2006) sugerem 0,50 como patamar aceitável.

`estat kmo`

CRIANDO NOVAS VARIÁVEIS

– Para criar novas variáveis automaticamente:

```
predict factor1 factor2
```

– Outra opção seria criar manualmente índices para cada conglomerado de variáveis.

– Por exemplo, se temos variáveis binárias:

```
gen factor1 = var1 + var2
```

– Por exemplo, se temos variáveis contínuas e que possuem valores mínimos e máximos compatíveis:

```
gen factor2 = (var3 + var4 + var5) / 3
```

predict factor1 factor2

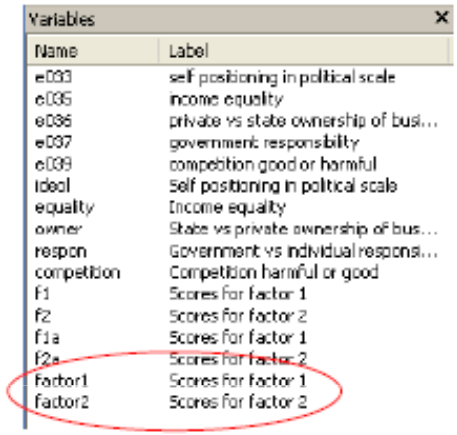
```
predict factor1 factor2 /*or whatever name you prefer to identify the factors*/
```

```
. predict factor1 factor2
(regression scoring assumed)

Scoring coefficients (method = regression; based on varimax rotated factors)
```

Variable	Factor1	Factor2
ideol	0.02868	0.42832
equality	-0.12258	0.53541
owner	0.58610	-0.05873
respon	0.07591	0.48119
competition	0.57225	-0.00014

These are the regression coefficients used to estimate the individual scores (per case/row)



Another option could be to create indexes out of each cluster of variables. For example, 'owner' and 'competition' define one factor. You could aggregate these two to create a new variable to measure 'market oriented attitudes'. On the other hand you could aggregate 'ideol', 'equality' and 'respon' to create an index to measure 'egalitarian attitudes'. Since all variables are in the same valence (liberal for small values, capitalist for larger values), we can create the two new variables as

```
gen market = (owner + competition)/2
gen egalitatiran = (ideol + equality + respon)/3
```

Fonte: Torres-Reyna, Oscar. s.d. *Getting Started in Factor Analysis (using Stata).* (http://dss.princeton.edu/training/)

SUGESTÕES DE LEITURA

Hamilton, Lawrence C. 2006. *Statistics with STATA (updated for version 9)*. Thomson Books/Cole.

Moraes, Odair Barbosa; Alex Kenya. 2006. *Utilização da Análise Fatorial para a Identificação de Estruturas de Interdependência de Variáveis em Estudos de Avaliação Pós-Ocupação*. XI Encontro Nacional de Tecnologia no Ambiente Construído (ENTAC).

Kim, Jae-on; Charles W. Mueller. 1978. *Factor Analysis. Statistical Methods and Practical Issues*. Sage publications.

Kim, Jae-on; Charles W. Mueller. 1978. *Introduction to Factor Analysis. What it is and How To Do It*. Sage publications.

StatNotes (http://faculty.chass.ncsu.edu/garson/PA765/factor.htm).

StatSoft (http://www.statsoft.com/textbook/stfacan.html).

Torres-Reyna, Oscar. s.d. *Getting Started in Factor Analysis (using Stata).* (http://dss.princeton.edu/training/)

Triola, Mario F. 2008. *Introdução à estatística*. 10ª ed. Rio de Janeiro: LTC. Cap.10.

UCLA (http://www.ats.ucla.edu/stat/stata/output/fa_output.htm).

Vincent, Jack. 1971. *Factor Analysis in International Relations. Interpretation, Problem Areas and Application*. University of Florida Press, Gainsville.