

A

Aceitação por amostragem (Acceptance sampling)

É um problema de controlo estatístico de qualidade e implica testar amostras aleatórias de um produto existente e decidir, sobre a aceitação de todo o lote, com base na qualidade da amostra.

A aceitação por amostragem é feita através de um plano de AMOSTRAGEM.

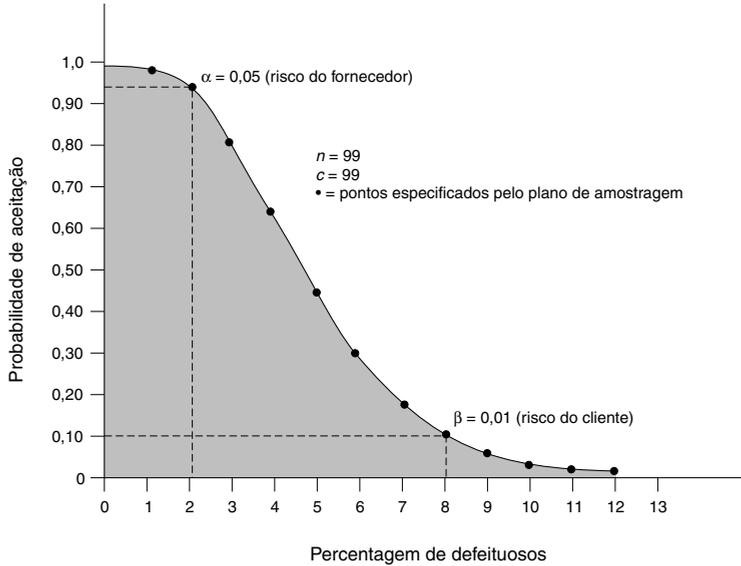
Um plano simples de amostragem é definido por n e C , sendo n o número de unidades na amostra e C o número de aceitação que indica o número máximo de artigos defeituosos que podem ser encontrados na amostra antes de o lote ser rejeitado.

O objetivo do fornecedor é assegurar que o plano de amostragem tenha uma baixa probabilidade de rejeitar os produtos bons. Os lotes são classificados de bons quando não contêm mais do que um determinado valor de defeituosos, chamado nível de qualidade aceitável (NQA). O objetivo do cliente é assegurar que o plano de amostragem tenha uma baixa probabilidade de aceitar lotes maus, isto é, que a percentagem de defeituosos seja maior do que um determinado valor (a percentagem de defeituosos tolerados no lote, PDDL). A probabilidade associada à rejeição de um lote bom é o risco do fornecedor (α). A probabilidade associada à aceitação de um lote mau é o chamado risco de cliente (β).

As curvas características operacionais podem ser calculadas partindo de uma distribuição binomial ou de Poisson.

Admitindo $n = 99$ e $C = 4$, o gráfico da probabilidade de aceitação vs percentagem de defeituosos, (curva característica operacional) para $NQA = 0,02$, $\alpha = 0,05$. $PDDL = 0,08$ e $\beta = 0,10$, está representado a seguir.

Nota-se que esta curva passa pelos pontos (NQA; $1 - \alpha$) e (PDDL; β).



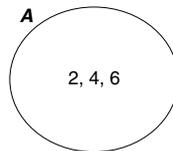
Acontecimento (Event)

Se lançarmos um dado sobre uma mesa e tomarmos nota da face voltada para cima, veremos que esta face pode ser um dos elementos do conjunto $\{1, 2, 3, 4, 5, 6\}$ que é o conjunto dos resultados possíveis.

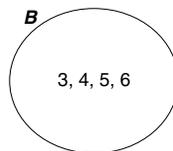
Considerando apenas o caso da saída da face par, isto é, a saída das faces 2, 4, 6, chamamos a este caso um acontecimento, e que é um subconjunto do conjunto dos resultados possíveis.

Chamando A ao acontecimento «saída de face par» e B ao acontecimento «saída de face superior a 2», isto é, saída de $\{3, 4, 5, 6\}$, podemos considerar o acontecimento da saída $\{2, 3, 4, 5, 6\}$ como ACONTECIMENTO UNIÃO de A com B que se indica por $A \cup B$ e se define como o acontecimento que ocorre se pelo menos um dos acontecimentos A ou B ocorrer.

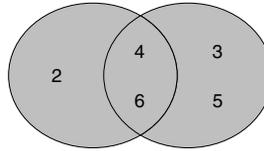
É costume representar um conjunto por um círculo (ou elipse). Assim, o acontecimento A será representado por:



e o acontecimento B , por



Então, o acontecimento $A \cup B$ será representado por :



pois pode ocorrer A , B ou ambos e engloba a região sombreada.

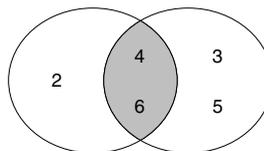
Partindo dos acontecimentos A e B considerados, A «saída da face par» e B «saída da face superior a 2» podemos definir o ACONTECIMENTO DIFERENÇA $A - B$, que é o acontecimento que ocorre se A ocorre, mas sem ocorrer B . Este acontecimento $A - B$ será apenas constituído pela saída da face 2.

Outro ACONTECIMENTO que se pode definir é o COMPLEMENTAR de um acontecimento A , ou ACONTECIMENTO CONTRÁRIO, representado por \bar{A} . Será constituída pelos resultados possíveis mas que não pertencem a A , neste caso portanto pela «saída da face impar», isto é, o sub-conjunto $\{1,3,5\}$.

Se chamarmos C ao acontecimento «saída de face impar» pode dizer-se que o acontecimento \bar{A} e o ACONTECIMENTO C são IDÊNTICOS, isto é, \bar{A} está contido em C ($A \subset C$) e C está contido em \bar{A} ($C \subset \bar{A}$) (há identidade de acontecimentos).

Se para os acontecimentos A e B procurássemos os elementos pertencentes ao mesmo tempo a A e a B , estávamos a considerar o ACONTECIMENTO INTERSECÇÃO de A com B , representado por $A \cap B$.

Em esquema, a intersecção de A com B será a região sombreada, compreendendo apenas as faces 4 e 6.



Um acontecimento que nunca pode ocorrer, como por exemplo «saída da face 7», é chamado ACONTECIMENTO IMPOSSÍVEL (conjunto vazio).

Se considerarmos o acontecimento A igual a «saída de face par» e o acontecimento C «saída de face impar», vê-se que não possuem elementos comuns, ou seja, a intersecção $A \cap B$ é o acontecimento impossível. Os acontecimentos nestas condições, isto é, com intersecção impossível, chamam-se ACONTECIMENTOS DISJUNTOS, MUTUAMENTE EXCLUSIVOS ou INCOMPATÍVEIS.

Acontecimento certo (Certain event)

É o acontecimento Ω , ESPAÇO DE RESULTADOS. De facto, qualquer que seja o acontecimento ω , ele é sempre pertencente a Ω .

Exemplo: Atirar um dado ao ar e sair um dos números 1, 2, 3, 4, 5 ou 6.

Ver \triangleright Espaço de resultados

Acontecimento complementar (*Complementary event*)

Ver ▷ Acontecimento

Acontecimento contrário (*Contrary event*)

Ver ▷ Acontecimento

Acontecimento diferença (*Difference event*)

Ver ▷ Acontecimento

Acontecimento impossível (*Impossible event*)

Ver ▷ Acontecimento

Acontecimento intersecção (*Intersection event*)

Ver ▷ Acontecimento

Acontecimento união (*Union event*)

Ver ▷ Acontecimento

Acontecimentos disjuntos (*Disjoined events*)

Ver ▷ Acontecimento

Acontecimentos idênticos (*Identical events*)

Ver ▷ Acontecimento

Acontecimentos incompatíveis (*Incompatible events*)

Ver ▷ Acontecimento

Acontecimentos independentes (*Independent events*)

Dois acontecimentos A e B , do mesmo espaço de resultados, dizem-se independentes se:

$$P(A \cap B) = P(A) \cdot P(B).$$

Alisamento exponencial (*Exponential smoothing*)

É um FILTRO LINEAR que se aplica às SUCESSÕES CRONOLÓGICAS. Tem a particularidade de não perder as informações mais recentes. É definido por: $y_t = \lambda_0 x_t + \lambda_1 x_{t-1} + \lambda_2 x_{t-2} + \dots + \lambda_m x_{t-m}$

Considerando um número infinito de desfasamentos e admitindo que os pesos diminuem de forma exponencial, isto é:

$$m = \infty \text{ e } \lambda_s = \alpha(1 - \alpha)^s, \quad s = 0, 1, 2, \dots \text{ e } 0 < \alpha < 1,$$

vem, após o tratamento matemático:

$$y_t = \alpha x_t + (1 - \alpha)y_{t-1}.$$

Trata-se de uma fórmula de recorrência que permite obter cada termo da sucessão transformada a partir do termo anterior e da observação mais recente. Para um valor de α próximo de 1 o alisamento é menor e quanto mais próximo estiver de zero, maior ele será.

Ver ► **Filtros lineares**

Alpha de Cronbach (*Cronbach's alpha*)

É uma medida muito utilizada em ANÁLISE DE COMPONENTES PRINCIPAIS para verificação da consistência interna de um grupo de variáveis. É pois uma medida de fidelidade de cada dimensão e do modelo no geral, já que ele é tanto melhor quanto maior for este índice.

Para o seu cálculo tem-se: $\alpha_i = \frac{m_w \sqrt{e_i} - m_w}{m_w \sqrt{e_i} - \sqrt{e_i}}$, onde m_w é o número de variáveis ponderadas na análise de componentes principais e e_i são valores da componente i (Maroco, J., 2003). O seu valor varia de 0 a 1.

- Se $\alpha > 0,9$ a consistência interna é muito boa;
- Se $0,8 < \alpha \leq 0,9$ a consistência interna é boa;
- Se $0,7 < \alpha \leq 0,8$ a consistência interna é razoável;
- Se $0,6 < \alpha \leq 0,7$ a consistência interna é fraca;
- Caso o alpha seja negativo, o modelo será inviabilizado.

Amostra (*Sample*)

Pretendendo-se estudar a altura dos portugueses, um caminho a seguir seria medir a altura de cada português. Como facilmente se compreende, tal não era acessível, como também seria um processo de custos muito elevados. Assim, mediante vários processos possíveis (ver AMOSTRAGEM) é preciso extrair da população um subconjunto finito, representativo desta população, a que se chama amostra. Nesta serão então estudadas as características pretendidas: a altura dos seus componentes. Ou seja, a amostra é um subconjunto extraído da população e desta representativa nas características em estudo. Nem sempre as amostras refletem a estrutura da população de onde foram retiradas o que poderá levar a conclusões erradas.

Ver ► **Amostragem**

Amostra aleatória (*Random sample*)

Sejam X_1, X_2, \dots, X_n VARIÁVEIS ALEATÓRIAS INDEPENDENTES E IDENTICAMENTE DISTRIBUÍDAS (i.i.d.) e F a respetiva FUNÇÃO DE DISTRIBUIÇÃO. Ao conjunto $\{X_1, X_2, \dots, X_n\}$ chama-se amostra aleatória de tamanho n pertencente à população com função de distribuição F . Ao conjunto $\{x_1, x_2, \dots, x_n\}$ chama-se realização da amostra aleatória.

A amostra aleatória tem função de distribuição conjunta dada por

$$F(x_1, x_2, \dots, x_n) = \prod_{i=1}^n F(x_i) = F(x_1)F(x_2), \dots, F(x_n)$$

que se designa por distribuição da amostra.

Ver ▷ Variável aleatória

Amostra ordenada (*Ordered samples*)

Considerando-se a amostra $x = (1,4; 0,6; 4,7; -1,3; 3; 4,4)$, a amostra ordenada será: $(-1,3; 0,6; 1,4; 3; 4,4; 4,7)$.

O mínimo é $x_{1;6} = -1,3$ e o máximo é $x_{6;6} = 4,7$.

Por exemplo a terceira estatística ordinal é $x_{3;6} = 1,4$

Assim para uma amostra original $x = (x_1, x_2, \dots, x_n)$ a amostra ordenada indica-se por $(x_{1;n}, x_{2;n}, \dots, x_{n;n})$.

Para uma amostra genérica de dimensão n , $x_{1;n}$ é o mínimo, $x_{n;n}$ é o máximo e $x_{k;n}$ é a K -ésima estatística ordinal.

Na notação seguida $x_{k;n}$ para estatísticas ordinais, n é a dimensão da amostra, K é o posto ou ordem (*rank*) e pressupondo-se a ordenação crescente.

Amostragem (*Sampling*)

É o processo de seleção de uma amostra a partir da população.

Amostragem aleatória estratificada (*Random stratified sampling*)

É um tipo de amostragem aleatória. Nesta amostragem a amostra final corresponde à soma de K subamostras homogêneas, obtidas aleatoriamente a partir de cada um dos K estratos ou subpopulações da população considerada e mutuamente exclusivos dentro da população.

Amostragem casual (*Casual sampling*)

Quando na seleção de AMOSTRAS ALEATÓRIAS as n variáveis aleatórias observadas, componentes do vetor (X_1, X_2, \dots, X_n) , são INDEPENDENTES E IDENTICAMENTE DISTRIBUÍDAS, diz-se que se trata de amostragem casual.

Amostragem com reposição (*Sampling with replacement*)

Quando num processo de amostragem se seleciona um elemento e o mesmo é repostado no conjunto antes de selecionar o elemento seguinte, dizemos que o processo de amostragem é com reposição.

Amostragem multietapas (*Multistage sampling*)

É um tipo de amostragem aleatória. Neste processo são utilizados mais do que um dos métodos de amostragem aleatória. É muito utilizado em ciências sociais.

Amostragem objetiva (*Objective sampling*)

É um tipo de amostragem aleatória, em que a amostra é constituída com determinado objetivo em mente, como por exemplo, assegurar determinado tipo de sujeitos na amostra.

Amostragem por conveniência (*Sampling by convenience*)

É um tipo de amostragem não aleatória. Neste processo de amostragem os elementos da amostra são selecionados acidentalmente ou por conveniência.

Amostragem por grupos ou conglomerados (*Clustered sampling*)

É um tipo de amostragem aleatória. Neste tipo de amostragem a população é dividida em conglomerados. Um número apropriado de conglomerados é selecionado aleatoriamente e todos os elementos desses conglomerados são medidos.

É um processo muito útil quando é necessário cobrir extensas zonas geográficas.

Amostragem sem reposição (*Sampling without replacement*)

Quando se seleciona um elemento e o mesmo não é repostado no conjunto, estamos em presença de um processo de amostragem sem reposição.

Amostragem simples (*Simple sampling*)

É um tipo de amostragem aleatória. Pode ser COM REPOSIÇÃO e SEM REPOSIÇÃO. Se há reposição, extraem-se n elementos da população de modo a que cada elemento da população apresente a mesma probabilidade de ser selecionado, isto é, $\frac{1}{N}$, sendo N a dimensão da população

e cada amostra de dimensão n apresenta a mesma probabilidade de ser escolhida, $\frac{1}{N^n}$.

Se não há reposição, a probabilidade de extração dos elementos da população vai aumentando e cada amostra de dimensão n apresenta a mesma probabilidade de ser escolhida, $\frac{1}{C_n^N}$. Para este tipo de amostragem pode recorrer-se a tabelas de números aleatórios.

Amostragem sistemática (*Systematic sampling*)

É um tipo de amostragem aleatória. Neste tipo de amostragem os N elementos da população são numerados aleatoriamente e depois, consoante a taxa de amostragem, seleciona-se um número de ordem $K = \frac{N}{n}$ (inverso da taxa de amostragem) que obriga à seleção dos elementos de ordem de um número inteiro entre 1 e K da população até obter a dimensão n da amostra.

Exemplo: Seja $N = 1000$ elementos e para uma taxa de amostragem de 20%, isto é, $\frac{n}{N} = 0,20$ ou seja $n = 200$. O K será $\frac{1000}{200} = 5$ e então seleciona-se ao acaso um número entre 1 e 5, por exemplo 2, recolhendo em seguida os elementos da população de ordem 5, começando no 2, isto é: 2, 2 + 5, 2 + 10, 2 + 15, ..., até se ter recolhido os 200 elementos.

Amostras correlacionadas (*Correlated sampling*)

São AMOSTRAS não independentes, tendo pois os seu elementos correlacionados.

Amostras dependentes (ou amostras correlacionadas) (*Dependent sampling*)

Ver ▷ Amostras correlacionadas

Amostras emparelhadas (*Paired samples*)

São aquelas em que os indivíduos que as constituem estão de alguma forma relacionados. Nos ESTUDOS LONGITUDINAIS utilizam-se AMOSTRAS de medições repetidas, como por exemplo, em estudos para avaliar a eficácia de determinado tratamento onde se efetuam análises antes e depois do mesmo ser aplicado. Formam-se deste modo duas amostras, a amostra «antes» e a amostra «depois» do tratamento, que se designam amostras emparelhadas.

Ver ▷ Amostras dependentes

Amostras independentes (*Independent samples*)

É lançada uma moeda ao ar e em seguida lança-se um dado. Qual a probabilidade de obter a face 1, sabendo que saiu a face cara voltada para cima? Seja A o acontecimento «sair face cara voltada para cima» e B o acontecimento «sair 1».

Empiricamente, sentimos que os ACONTECIMENTOS A e B são INDEPENDENTES, ou seja, a realização de um acontecimento não condiciona a realização de outro. Isto é traduzido em linguagem probabilística por $P(A \cap B) = P(A) \cdot P(B)$ ou seja a probabilidade da realização simultânea de A e B é igual ao produto das probabilidades de realização dos acontecimentos A e B , respetivamente.

Ver ▷ Probabilidade condicional

Amplitude da classe (*Class amplitude*)

Quando se trata de VARIÁVEIS CONTÍNUAS agrupam-se os dados em CLASSES. Na definição de classe admitem-se intervalos, geralmente fechados à esquerda e abertos à direita (por convenção)

Exemplos de classes:

[0; 300[

[300;600[

Como se vê cada intervalo da classe tem um limite inferior e um limite superior, e é precisamente a diferença entre esses limites que se chama amplitude de classe.

Para as classes acima indicadas a amplitude é constante:

$$300 - 0 = 600 - 300 = 300$$

De um modo geral a amplitude de j -ésima classe h_j é definida por:

$$h_j = l_j - l_{j-1}, \quad j = 1, 2, m$$

Ver ▷ Dados classificados

Amplitude do intervalo de variação (Range amplitude)

Ver ▷ Amplitude total

Amplitude interquartis (AIQ) (Interquartile range)

É uma MEDIDA DE DISPERSÃO. É a diferença entre o terceiro e o primeiro QUARTIL, isto é $A_{IQ} = Q_3 - Q_1$.

Este intervalo contém 50% das observações centrais da coleção de dados.

Para o conjunto de observações {1, 4, 5, 6, 8, 9, 11, 14, 15, 18, 20} vem: $Q_3 - Q_1 = 15 - 5 = 10$

Ver ▷ Dispersão; Quartis

Amplitude semi-interquartis (Semi-interquartile range)

É uma MEDIDA DE DISPERSÃO que é dada por $\frac{Q_3 - Q_1}{2}$, ou seja, é metade da AMPLITUDE INTERQUARTIS.

Ver ▷ Dispersão; Quartis

Amplitude total (R) (Range)

Também chamada amplitude do intervalo de variação e é a MEDIDA DE DISPERSÃO mais fácil de calcular:

$R = x_{\max} - x_{\min}$, isto é, a diferença entre o valor máximo e mínimo da variável

Considerem-se as sucessões de valores {0, 1, 2, 3, 96, 97, 98, 99, 100} e {0, 10, 20, 30, 40, 60, 80, 90, 100}

Qualquer delas tem a amplitude total igual a 100, mas a sua dispersão é muito diferente.

De facto a amplitude total tem a grande desvantagem de só ter em conta os valores extremos que a variável toma, e portanto, não ser sensível aos valores intermédios.

Ver ▷ Dispersão

Análise combinatória (Combinatory analysis)

É um conjunto de técnicas que permitem estudar a constituição de grupos formados com todos ou apenas alguns dos elementos dados, grupos estes que podem diferir uns dos outros, quer pela natureza dos elementos, quer pela ordem como estão dispostos.

Análise da potência de um teste (Power test analysis)

Ver ▷ Potência de um teste

Análise de *clusters* (*Clusters analysis*)

Os métodos de análise de *clusters*, são procedimentos de ESTATÍSTICA MULTIVARIADA que tentam organizar um conjunto de indivíduos, para os quais é conhecida informação detalhada, em grupos relativamente homogêneos (*clusters*).

É muito aplicada na Sociologia, Medicina, Psicologia, Economia e Gestão, particularmente no que toca a estudos de mercado.

Dado um conjunto de n indivíduos, para os quais existe informação sobre a forma de p variáveis, o método de análise *clusters* procede ao agrupamento dos indivíduos em função da informação existente, de tal modo que os indivíduos pertencentes ao mesmo grupo sejam tão semelhantes quanto possível e sempre mais semelhantes aos elementos do mesmo grupo do que a elementos dos restantes grupos.

Daí os nomes de método de partição, classificação ou taxonomia, pelos quais também é conhecida a análise de *clusters*.

Nesta análise não existe qualquer tipo de dependência entre as variáveis.

Compreende cinco etapas:

1. Seleção de indivíduos ou de uma amostra de indivíduos a serem agrupados.
2. Definição de um conjunto de variáveis a partir das quais será obtida a informação necessária ao agrupamento dos indivíduos.
3. Definição de uma MEDIDA DE SEMELHANÇA OU DISTÂNCIA entre cada dois dos indivíduos.
4. Escolha de um critério de agregação ou desagregação dos indivíduos.
5. Validação dos resultados.

Os *clusters* podem ser estudados por MÉTODOS HIERÁRQUICOS DE AGRUPAMENTO ou por MÉTODOS NÃO HIERÁRQUICOS DE AGRUPAMENTO, e para avaliar o número de *clusters* a reter pode recorrer-se a índices como por exemplo a distância entre *clusters* e o critério do R quadrado.

Ver ► **Métodos hierárquicos de agrupamento; Métodos não hierárquicos de agrupamento; Medidas de semelhança e dissemelhança**

Análise de componentes principais (ACP) (*Principal components analysis*)

É uma técnica multivariada que transforma um conjunto de variáveis correlacionadas num conjunto menor de variáveis independentes, combinações lineares das variáveis originais, chamadas componentes principais.

Por exemplo para m componentes e p variáveis ($m \leq p$), tem-se:

$$\begin{aligned} CP_1 &= a_{11} X_1 + a_{21} X_2 + \dots + a_{p1} X_p \\ CP_2 &= a_{12} X_1 + a_{22} X_2 + \dots + a_{p2} X_p \\ &\dots \\ CP_m &= a_{1m} X_1 + a_{2m} X_2 + \dots + a_{pm} X_p \end{aligned}$$

onde a_{ij} é o peso da variável X_i na componente principal CP_j .

A primeira componente principal explica a maior proporção da variância nas variáveis originais.

A componente seguinte explica a maior proporção da variância não explicada pela primeira componente e assim sucessivamente.

A análise de componentes principais pode ser feita com variáveis estandardizadas, isto é, variáveis $Z = \frac{X - \bar{X}}{s}$. Neste caso a nova MATRIZ DE VARIÂNCIAS-COVARIÂNCIAS não é mais do que a MATRIZ DE CORRELAÇÃO AMOSTRAL \mathbf{R} e as componentes principais passam a ser os vetores próprios associados aos valores próprios de \mathbf{R} .

Exemplo:

Supondo-se que duas variáveis critério Y_1 e Y_2 têm médias e variâncias iguais a:

$$\bar{Y}_1 = 45$$

$$\bar{Y}_2 = 25$$

$$s_{11} = 25 \quad s_{22} = 9$$

e uma matriz de correlações igual a $\mathbf{R} = \begin{bmatrix} 1,0 & 0,8 \\ 0,8 & 1,0 \end{bmatrix}$, podem obter-se as componentes principais a partir da matriz de correlações, o que significa que se utilizaram os dados estandardizados.

Assim, as componentes principais tomam a forma:

$$CP_1 = a_{11} \frac{Y_1 - \bar{Y}_1}{s_1} + a_{21} \frac{Y_2 - \bar{Y}_2}{s_2}$$

$$CP_2 = a_{12} \frac{Y_1 - \bar{Y}_1}{s_1} + a_{22} \frac{Y_2 - \bar{Y}_2}{s_2}$$

Os valores próprios de \mathbf{R} são as soluções da equação característica $(1 - \lambda)^2 - 0,8^2 = 0$, donde $\lambda_1 = 1,8$ e $\lambda_2 = 0,2$.

As contribuições para a explicação da variância dos dados da primeira e segunda componente são respetivamente:

$$\frac{1,8}{1,8 + 0,2} = 90\% \quad \text{e} \quad \frac{0,2}{1,8 + 0,2} = 10\%$$

O cálculo dos vetores próprios, com norma unitária, associados a $\lambda_1 = 1,8$ e $\lambda_2 = 0,2$ fornece os valores: $a_{11} = 0,707$, $a_{21} = 0,707$, $a_{12} = 0,707$ e $a_{22} = -0,707$.

Assim, as componentes principais são dadas pelas expressões:

$$CP_1 = 0,707 \left(\frac{Y_1 - 45}{5} \right) + 0,707 \left(\frac{Y_2 - 25}{3} \right) = 0,141 Y_1 + 0,236 Y_2 - 12,255$$

$$CP_2 = 0,707 \left(\frac{Y_1 - 45}{5} \right) - 0,707 \left(\frac{Y_2 - 25}{3} \right) = 0,141 Y_1 - 0,236 Y_2 - 0,471$$

**Ver ▶ Teste de Kaiser – Meyer – Olkin (KMO); Teste de esfericidade de Bartlett;
Rotações das componentes principais; Scree-plot**

Análise de covariância (ANCOVA) (*Analysis of covariance*)

É um método que combina a ANÁLISE DE VARIÂNCIA com a de REGRESSÃO, incluindo além dos pressupostos de INDEPENDÊNCIA, NORMALIDADE e HOMOCEDESTICIDADE, os seguintes:

As retas de regressão para cada grupo assumem-se ser paralelas, isto é, a regressão de Y em X é a mesma para cada grupo;

Existe uma ASSOCIAÇÃO LINEAR entre a variável concomitante (X) e a variável dependente (Y);

As VARIÁVEIS CONCOMITANTES são fixas e não contêm erros de medição;

A variável concomitante é uma variável independente introduzida, de natureza quantitativa, X, que se assume estar correlacionada com a variável dependente, mas não com o fator. Permite diminuir a variância não explicada entre os grupos e, portanto, reduzir o erro do modelo.

Análise de estudos de caso-controlo (*Case-control studies analysis*)

As frequências observadas num ESTUDO DE CASO-CONTROLO podem dispor-se numa tabela do tipo:

		Exposição ao fator		
		Sim	Não	Total
Estado da doença	Caso	a	b	a + b
	Controlo	c	d	c + d
	total	a + c	b + d	n = a + b + c + d

Nestes estudos, a aproximação à medida RISCO RELATIVO calcula-se a partir das FREQUÊNCIAS RELATIVAS da exposição dos casos e dos controlos. O ODD-RATIO permite obter uma medida de ASSOCIAÇÃO que, sendo a doença rara, se pode considerar semelhante ao risco relativo.

O *odd-ratio* é determinada por:

$$OR = \frac{\frac{\frac{a}{a+c}}{\frac{c}{a+c}}}{\frac{\frac{b}{b+d}}{\frac{d}{b+d}}} = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{ad}{bc}$$

Ver ► Odds-ratios