

# Aplicação de método e técnica multivariados para previsão de variáveis termoambientais e perceptivas

ANAND SUBRAMANIAN

ANTONIO SOUTO COUTINHO

LUIZ BUENO DA SILVA

UFPB

## Resumo

Um ambiente ergonômico ideal para exercer-se atividades bancárias precisa estar termicamente adequado, principalmente segundo a percepção térmica daqueles que realmente promovem a produtividade: os seres humanos. Esta pesquisa possibilitou conhecer-se as características termoambientais destas atividades e, a partir dos dados coletados, aplicou-se o método da Regressão Linear e a técnica exploratória Análise Discriminante para determinar quais variáveis térmicas representam melhor a sensação térmica declarada pelo trabalhador. Para isso, determinaram-se as variáveis do balanço térmico e a sensação térmica de cada trabalhador, baseando-se em normas internacionais. Constatou-se que a temperatura de bulbo seco foi a variável que representou melhor estas preferências e, em situação de conforto, 23,79°C é a temperatura ideal para que estas atividades sejam exercidas com satisfação térmica.

## Palavras-chave

Sensação térmica, variáveis térmicas, ergonomia, regressão linear, análise discriminante.

## *Application of multivariate method and technique for prediction of thermo-environmental and perceptive variables*

## Abstract

*An ergonomically ideal environment for bank services needs to be thermally adequate, mainly according to the thermal perceptions of those responsible for the productivity, namely, human beings. This research has made it possible to recognize the relevant thermo-environmental characteristics of these activities and, based on the collected data, to apply the Linear Regression method and Discriminant Analysis exploratory technique to determine the thermal variables that may best represent the employee's reported thermal sensations. For this purpose, the variables of thermal balance and sensation of each worker were determined, based on available international standards. It was found that the dry bulb temperature was the one that better represented these preferences. In the comfort situation, 23.79°C was the ideal temperature for these activities to be performed with thermal satisfaction.*

## Key words

*Thermal sensation, thermal variables, ergonomics, linear regression, discriminant analysis.*

## INTRODUÇÃO

Desde o início de sua existência, o ser humano tem procurado se proteger das intempéries e buscado as melhores condições de conforto térmico. Partindo da caverna e da fogueira, chegou aos atuais sistemas de ventilação e ar condicionado atuados por controle remoto. Com o tempo, verificou que tinha mais motivação para as tarefas diárias quando o ambiente lhe proporcionava conforto; e assim viu a possibilidade de aumentar a produtividade melhorando as condições de conforto. Todavia, o conforto é função de quatro variáveis climáticas e duas pessoais que atuam simultaneamente, tornando a pesquisa bastante árdua.

Tem-se constatado a necessidade de se estabelecer relações que possibilitem prever uma ou mais variáveis em termos de outras, ou quando se quer saber a relação entre essas variáveis. Por exemplo, as sensações térmicas declaradas subjetivamente pelos trabalhadores possuem uma forte ou fraca associação com o percentual de insatisfeitos calculados através de modelos desenvolvidos em outros países? Existem uma ou mais variáveis do balanço térmico que explicariam melhor que as outras a variação da sensação térmica declarada por trabalhadores durante uma jornada de trabalho? A relação entre os *valores teóricos de sensação térmica* e a *percentagem de insatisfeitos* (índices da norma 7730/94) pode ser ou não mais representativa do que os correspondentes valores declarados subjetivamente?

Um problema de grande importância no estudo de variáveis térmicas que possuem distribuição normal é a análise simultânea entre elas, com a finalidade de averiguar se existe alguma correlação significativa entre as mesmas ou de investigar a possibilidade de se fazer previsões a respeito de valores de uma das variáveis correlacionadas, com base no conhecimento dos valores das outras.

Como se poderia melhor controlar alguma variável que compõe o ambiente de trabalho, entre essas, as de conforto térmico e as que fazem parte do processo produtivo, sem antes fazer um minucioso estudo do grau de relação entre essas variáveis? Como descobrir os problemas provenientes da irritabilidade ocular ou do ressecamento da mucosa nasal em profissionais que exercem suas atividades em ambientes informatizados climatizados sem, primeiro, verificar a relação entre a variação da umidade do ar e o tempo da atividade realizada pelos mesmos durante várias jornadas de trabalho?

Assim, se o conhecimento prévio sobre a atividade, as características do ambiente de trabalho, o tempo de permanência do homem nesse ambiente, a realização do planejamento em conjunto de projetos lumínico e térmicos, e se as particularidades do próprio homem podem contribuir para o desenvolvimento de projetos ergonômicos de ambien-

tes de trabalho seguros e produtivos, métodos e técnicas multivariados poderão auxiliar na previsão e determinação de variáveis e valores ideais de conforto térmico, como também avaliar se estas variáveis e estes valores estão de acordo com as preferências térmicas de trabalhadores, o que contribuiria de maneira significativa para o sucesso de projetos ergonômicos de ambientes de trabalho. Desta forma, este artigo apresenta teoricamente os modelos de regressão linear seguidos da análise discriminante, com o objetivo de prever, determinar e classificar variáveis termoambientais segundo a percepção térmica das pessoas que exercem suas atividades em um centro de processamento de dados do setor bancário.

## CONFORTO TÉRMICO

Pode-se considerar o homem como um sistema térmico e aplicar a primeira lei da termodinâmica para analisar as parcelas de energia térmica envolvidas na interação desse sistema com o ambiente no qual está envolvido. Essas parcelas são: metabolismo, trabalho externo, convecção e evaporação que ocorrem no aparelho respiratório e na pele, assim como a radiação também na pele. Os processos realizados na pele sofrem a influência da vestimenta utilizada pela pessoa. A interação térmica homem-ambiente envolve quatro variáveis climáticas: temperatura, umidade e velocidade do ar e temperatura radiante média; e duas pessoais: metabolismo, função da atividade desenvolvida, e resistência térmica das vestes.

No balanço térmico da pessoa o saldo deve ser nulo, pois o positivo implicaria o surgimento de doenças típicas de ambientes quentes, como tontura e desfalecimento, desidratação, distúrbios cutâneos, psicose, hipertermia, etc.; por outro lado, o saldo negativo pode provocar os males característicos dos ambientes frios: enregelamento dos membros, ulcerações, crises reumáticas e respiratórias e até a hipotermia. Todavia, dentro de certos limites, o *sistema de termorregulação* anula o saldo através de mecanismos tais como vasodilatação, vasoconstrição, sudorese e tiritar. Esses mecanismos exigem esforço do organismo, implicando desconforto. Entre os limites das faixas  *muito frio* e  *muito quente*, há um pequeno intervalo que caracteriza o *conforto térmico*, no qual o esforço do sistema de termorregulação é mínimo.

## REGRESSÃO LINEAR

A análise de regressão consiste em um método de modelagem que avalia a relação entre uma variável dependente contínua  $Y$  e uma ou mais variáveis contínuas independentes  $X_1, X_2, \dots, X_k$ . O objetivo da análise de regressão é identificar a função que descreve, da melhor forma, a relação entre essas

variáveis para que se possa prever que valor a variável dependente vai assumir quando forem atribuídos valores para a variável independente (RAGSDALE, 2001).

**O modelo de regressão linear simples (MRLS)**

Conforme Charnet *et al.* (1999), seja  $Y$  uma variável aleatória de interesse, muitas vezes denominada variável resposta, e  $X$  uma variável aleatória denominada auxiliar ou regressora. O MRLS descreve a variável  $Y$  como uma soma de uma quantidade determinística e uma quantidade aleatória. A parte determinística, uma reta em função de  $X$ , representa a informação sobre  $Y$  que já se pode “esperar”, apenas com o conhecimento da variável  $X$ . A parte aleatória, denominada erro, representa os inúmeros fatores que, conjuntamente, podem interferir em  $Y$ . Assim, pode-se interpretar que o erro provoca uma distorção sobre a parte determinística de  $Y$ .

Suponha que erros positivos ou negativos possam ocorrer, mais que isso, e que a variável erro tenha esperança igual a zero. Suponha, ainda, que a variância da variável erro não depende do valor específico de  $X$ . Utilizando-se  $\beta_0$  e  $\beta_1$  para denotar os coeficientes da reta;  $\varepsilon$ , a variável erro;  $\sigma^2$ , a variância da variável erro; e  $x_p$ , um valor específico da variável  $X$ , então o MRLS pode ser expresso como

$$y_i = \beta_0 + \beta_1 \cdot x_i + \varepsilon_i \quad (1)$$

com  $\beta_0$ ,  $\beta_1$  e  $x_i$  constantes;  $E[\varepsilon_i] = 0$ ;  $Var[\varepsilon_i] = \sigma^2$ ;  $Cov[\varepsilon_p, \varepsilon_j] = 0$ , para  $i \neq j$  e  $i, j = 1, \dots, n$ . Quando a condição do modelo de probabilidade do erro é a normalidade, o modelo de regressão linear simples amostral correspondente adota a equação (4) sujeita às seguintes restrições:  $\beta_0$ ,  $\beta_1$  e  $x_i$  constantes;  $\varepsilon_i \sim N(0; \sigma^2)$ ;  $Cov[\varepsilon_p, \varepsilon_j] = 0$  para  $i \neq j$  e  $i, j = 1, \dots, n$ .

**Estimadores do MRLS**

Para Magalhães e Lima (2002) a estimação de  $\beta_0$  e  $\beta_1$  pode ser feita através do método dos mínimos quadrados, que consiste em minimizar a soma dos quadrados dos resíduos, obtidos através da diferença entre valores observados  $y$  e valores esperados  $E(Y | X = x)$ , calculados para cada  $X = x$ .

Como em geral os pontos não estão perfeitamente alinhados, escolhe-se a “melhor” reta possível no sentido de minimizar a Soma de Quadrados  $SQ(\beta_0, \beta_1)$ , dada por:

$$SQ(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - E(Y_i | X_i = x_i))^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 \cdot x_i)^2 = \sum_{i=1}^n e_i^2 \quad (2)$$

A solução do sistema de equações envolvendo as derivadas de  $SQ(\beta_0, \beta_1)$ , em relação a  $\beta_0$  e  $\beta_1$  fornecerá os estimado-

res de mínimos quadrados para  $\beta_0$  e  $\beta_1$ . Verifica-se então:

$$b_0 = \frac{(\sum y) \cdot (\sum x^2) - (\sum x) \cdot (\sum x \cdot y)}{n \cdot (\sum x^2) - \sum (x)^2} \quad (3)$$

$$b_1 = \frac{n \cdot \sum x \cdot y - (\sum x) \cdot (\sum y)}{n \cdot (\sum x^2) - \sum (x)^2} \quad (4)$$

Anderson *et al.* (2002) definem a equação da regressão linear simples estimada da seguinte maneira:

$$\hat{y} = b_0 + b_1 \cdot x \quad (5)$$

**Medidas de Variação na Regressão**

De acordo com Levine *et al.* (2000), para examinar como a variável independente prevê bem a variável dependente em um modelo estatístico, é preciso desenvolver diversas medidas de variação. A seguir serão expostas algumas dessas medidas.

**Obtenção da soma dos quadrados mínimos**

A primeira medida, a Soma Total dos Quadrados,  $SQT$ , é uma medida de variação dos valores de  $Y_i$  em torno de sua média aritmética  $\bar{Y}$ . Em uma análise de regressão, a soma total dos quadrados pode ser subdividida em variações explicadas ou Soma dos Quadrados devida à Regressão,  $SQReg$ , que é atribuída à relação entre  $X$  e  $Y$ , e variações inexplicadas ou a Soma de Quadrados dos Resíduos,  $SQR$ , que é atribuída a outros fatores diferentes da relação entre  $X$  e  $Y$ .

Essas medidas de variação podem ser representadas conforme demonstração abaixo:

Suponha que

$$SQReg = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 = b_0 \sum_{i=1}^n Y_i + b_1 \sum_{i=1}^n X_i Y_i - n\bar{Y}^2 \quad (6),$$

$$SQR = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n Y_i^2 - b_0 \sum_{i=1}^n Y_i - b_1 \sum_{i=1}^n X_i Y_i \quad (7),$$

assim

$$SQT = b_0 \sum_{i=1}^n Y_i + b_1 \sum_{i=1}^n X_i Y_i - n\bar{Y}^2 + \sum_{i=1}^n Y_i^2 - b_0 \sum_{i=1}^n Y_i - b_1 \sum_{i=1}^n X_i Y_i,$$

$$SQT = \sum_{i=1}^n Y_i^2 - n\bar{Y}^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (8), \text{ portanto}$$

$$SQT = SQReg + SQR \quad (9)$$

**O Coeficiente de Determinação**

Suponha que  $SQR$ ,  $SQReg$  e  $STQ$  estejam bem definidas, então o coeficiente de determinação  $r^2$  é expresso, segundo Levine *et al.* (2000) como:

$$r^2 = \frac{SQReg}{STQ} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (10)$$

Charnet *et al.* (1999) ainda esclarecem que o coeficiente de determinação  $r^2$  é a proporção da variabilidade dos  $Y$ s observados explicada pelo modelo considerado em que o valor de  $r^2$  pertence ao intervalo  $[0;1]$ ; e quanto mais próximo de 1, melhor o ajuste do modelo considerado.

### O coeficiente de correlação

De acordo com Magalhães e Lima (2002), o coeficiente de correlação é o coeficiente entre a covariância e o produto dos desvios-padrão de  $X$  e  $Y$ , isto é:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \quad (11)$$

A divisão pelo produto dos desvios-padrão,  $\sigma_X \sigma_Y$ , tem a função de padronizar a medida e torná-la possível de ser utilizada para comparações com outras variáveis. Para valores próximos de  $\pm 1$ , a correlação é considerada forte.

Conforme Levine *et al.* (2000), se uma análise de correlação está sendo realizada em um conjunto de dados, o coeficiente de correlação da amostra  $r$  pode ser calculado através da equação (12):

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \cdot \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (12)$$

### Pressupostos do MRLS

Os quatro principais pressupostos da regressão segundo Levine *et al.* (2000) são:

**Normalidade** – requer que os valores de  $Y$  sejam normalmente distribuídos para cada valor de  $X$ . Enquanto a distribuição dos valores de  $Y_i$  em torno de cada nível de  $X$  não for extremamente diferente de uma distribuição normal, inferências sobre a linha de regressão e sobre os coeficientes de regressão não serão seriamente afetadas.

**Homocedasticidade** – requer que as variações em torno da linha de regressão sejam constantes para todos os valores de  $X$ . Isto significa que  $Y$  varia na mesma proporção, quando  $X$  for um valor baixo e quando  $X$  for um valor elevado.

**Independência de erros** – requer que o erro (diferença residual entre valores observados e previstos de  $Y$ ) deva ser independente para cada valor de  $X$ .

**Linearidade** – esse pressuposto estabelece que a relação

entre variáveis deve ser linear. Duas variáveis poderiam perfeitamente ser relacionadas de uma maneira não-linear, e o coeficiente de correlação linear seria igual a 0, indicando não haver qualquer relação.

### Teste da falta de ajuste do MRLS

Charnet *et al.* (1999) enfatizam que para testar estatisticamente a falta de ajuste do MRLS, precisa-se de replicações, ou seja, deve-se ter pelo menos dois valores da variável resposta para alguns valores da variável regressora.

Suponha que se tenha  $k$  valores distintos da variável regressora  $X: x_1, \dots, x_k$ . Para o valor de  $x_i$  observa-se  $n_i$  valores da variável resposta  $Y$ , donde observações. Neste contexto usa-se a notação  $y_{ij}$ , sendo o índice  $i$  correspondente ao índice de  $x_i$  e o índice  $j$  uma enumeração dos  $n_i$  valores. Assim, tem-se que

$$\begin{aligned} x_1 &\rightarrow y_{11}, y_{12}, \dots, y_{1n_1} \\ x_2 &\rightarrow y_{21}, y_{22}, \dots, y_{2n_2} \\ &\vdots \\ x_k &\rightarrow y_{k1}, y_{k2}, \dots, y_{kn_k} \end{aligned}$$

As expressões acima apresentam um estimador natural para  $\text{Var}[Y|x]$ ,  $\sigma^2$ , que não está baseado no MRLS. Por exemplo, se observarmos  $y_{11}, y_{12}, \dots, y_{1n_1}$  quando  $X = x_1$ , um estimador para  $\text{Var}[Y|x]$  seria dado pela variância amostral desta amostra, ou seja:

$$\frac{1}{n-1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \quad (13)$$

sendo  $\bar{y}_i$  a média amostral de todos os valores observados para  $X = x_i$ . Pode-se definir outros estimadores através das variâncias amostrais relativas aos outros valores de  $X$ . No entanto é melhor combinar as  $k$  variâncias amostrais utilizando assim toda a amostra.

Denota-se o estimador  $\hat{\sigma}^2_{ep}$ , no qual o subscrito *ep* significa “erro puro”, por depender da variação pura e não do MRLS, como

$$\hat{\sigma}^2_{ep} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{\sum_{j=1}^k (n_i - 1)} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{n - k} \quad (14)$$

sendo  $\bar{y}_i$  a média amostral dos valores de  $Y$  correspondentes a  $X = x_i$ . No contexto do MRLS tem-se que o estimador de  $\text{VAR}[Y|x]$  é  $\hat{\sigma}^2 = \frac{SQR}{(n-2)}$ , que depende da adequação do modelo. Através da comparação destes dois estimadores de  $\sigma^2$ , pode-se verificar a adequação do MRLS. Usando a notação acima, a equação do MRLS amostral é expressa da seguinte forma:

$$y_{ij} = \beta_0 + \beta_1 x_i + \varepsilon_{ij}, \quad j = 1, \dots, n_i, \quad i = 1, \dots, k \quad (15)$$

e a Soma de Quadrados dos Resíduos é dada por:

$$SQR = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \hat{y}_i)^2 \quad (16)$$

Nota-se que para cada valor distinto  $x_i$  tem-se apenas um valor  $y_{ij}$ , não sendo necessário o segundo índice. Neste caso a  $SQR$  pode ser particionada em duas outras somas de quadrados, conforme é demonstrado abaixo:

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \hat{y}_i)^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^k n_i (\hat{y}_i - \bar{y}_i)^2 \quad (17)$$

As parcelas da expressão dada na equação (17) têm uma denominação especial:

$\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \hat{y}_i)^2$  é a Soma de quadrados dos Resíduos –  $SQR$ , representando a variação em torno da reta;

$\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$  é a soma de quadrados do erro puro –  $SQE_p$ , representando a variação de  $Y$ , para  $X$  fixo, independente do modelo;

$\sum_{i=1}^k n_i (\hat{y}_i - \bar{y}_i)^2$  é a soma de quadrados da falta de ajuste –  $SQF_a$ , representando a falta de ajuste do modelo.

Através da equação (17), é estabelecido que  $SQR = SQE_p + SQF_a$ . Note que a  $SQR$  na equação (14) é o numerador da expressão para  $\hat{\sigma}^2$ , e a  $SQE_p$  indica diferenças entre os dois estimadores e, conseqüentemente, demonstra evidências contra a adequação do MRLS. Outra interpretação é a de que quanto mais alto for o valor da  $SQF_a$  pior será o ajuste do modelo, visto que se o ajuste fosse perfeito  $\hat{y}_i$  seria exatamente igual a  $\bar{y}_i$ , donde  $SQF_a = 0$ , para  $i = 1, \dots, k$ .

Considerando os argumentos acima, pode-se construir um teste para as hipóteses:

- $H_0$  : O MRLS é adequado
- $H_1$  : O MRLS não é adequado

com base na estatística  $\frac{SQF_a / (k-2)}{SQE_p / (n-k)}$  a qual, sob hipótese  $H_0$ , tem distribuição  $F$  de Snedecor com  $(k-2)$  e  $(n-k)$  graus de liberdade (Teorema de Cochran). Portanto, rejeita-se a hipótese de o modelo ser adequado se a estatística do teste for maior que  $F_{k-2, n-k}(\alpha)$ , o quantil  $(1-\alpha)$  de  $F$  com  $(k-2)$  e  $(n-k)$  graus de liberdade, para algum nível de significância prefixado.

### Teste para os parâmetros da população no MRLS

De acordo com Levine *et al.* (2000), pode-se determinar se existe relação significativa entre as variáveis  $X$  e  $Y$  testan-

do se  $\beta_1$  (a verdadeira inclinação) é igual a 0. Se essa hipótese for rejeitada, poder-se-ia concluir que há evidências de uma relação linear. As hipóteses nula e alternativa poderiam ser declaradas da seguinte maneira:

- $H_0$ :  $\beta_1 = 0$  (não existe relação)
- $H_1$ :  $\beta_1 \neq 0$  (existe relação)

A estatística de teste  $t$  é igual à diferença entre a inclinação da população dividida pelo erro padrão da inclinação, conforme apresentado na equação (18).

$$t = \frac{b_1 - \beta_1}{S_{b_1}} \quad (18)$$

onde

$$S_{b_1} = \frac{S_{YX}}{\sqrt{\sum_{i=1}^n X_i^2 - n\bar{X}}} \quad (19)$$

e  $S_{YX}$  é o erro padrão da estimativa<sup>1</sup>.

Este teste segue uma distribuição  $t$  com  $n-2$  graus de liberdade. A hipótese de o modelo ser adequado é rejeitada se a estatística do teste for maior que o  $t$  tabelado. Vale salientar que sob a hipótese nula,  $\beta_1 = 0$ , de modo que

$$t = \frac{b_1}{S_{b_1}} \quad (20)$$

### Análise de Resíduos

Para Charnet *et al.* (1999), a análise de um modelo de regressão linear tem uma relação muito forte com a qualidade do ajuste feito, bem como com a confiabilidade dos testes estatísticos sobre os parâmetros do modelo. Nesse sentido, a análise dos resíduos tem uma importância fundamental na verificação da qualidade dos ajustes de modelos. Basicamente, essa análise fornece evidências sobre possíveis violações nas suposições do modelo, tais como a de normalidade e a homocedasticidade, e quando for o caso ainda fornece indícios da falta de ajuste do modelo proposto.

Para Levine *et al.* (2000), o resíduo é igual ao valor observado  $Y_i$  menos o valor previsto  $\hat{Y}_i$ , isto é,

$$e_i = Y_i - \hat{Y}_i \quad (21)$$

Pode-se avaliar a adequação do modelo de regressão ajustado plotando os resíduos no eixo vertical e os valores correspondentes aos valores de  $X_i$  da variável independente no eixo horizontal. Se o modelo ajustado for apropriado para os dados, não haverá padrão aparente de resíduos em relação a  $X_i$ . No entanto, se o modelo ajustado não for apropriado, existirá uma relação entre os valores de  $X_i$  e os resíduos  $e_i$ .

É possível calcular, também, os resíduos padronizados e

os de Student. Os resíduos de Student, expressos na equação (22) são os padronizados, ajustados para a distância do valor de  $X$  médio, isto é,

$$RS_i = \frac{e_i}{S_{YX}\sqrt{1-h_i}} \quad (22)$$

$$\text{onde } h_i = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

Esses resíduos de Student permitem considerar suas magnitudes em unidades que refletem as variações padronizadas em torno da linha de regressão.

### Medindo a autocorrelação

Um dos pressupostos do modelo de regressão básico, já citado anteriormente, é a independência dos resíduos. Esse pressuposto é geralmente violado quando os dados são coletados ao longo de períodos seqüenciais de tempo, uma vez que um resíduo, em qualquer ponto no tempo, pode tender a ser idêntico a resíduos em pontos adjacentes no tempo. Levine *et al.* (2000) denominam este padrão presente nos resíduos de autocorrelação. Quando uma substancial correlação se encontra presente em um conjunto de dados, a validade de um modelo de regressão ajustado pode ficar seriamente comprometida.

O método mais simples para averiguar se existe autocorrelação em um conjunto de dados é plotando os resíduos ou os padronizados, em ordem cronológica. Se o efeito de uma autocorrelação positiva estiver presente, grupos de resíduos com o mesmo sinal estarão presentes e um padrão aparente será prontamente detectado.

Além dos gráficos de resíduos, a autocorrelação também pode ser detectada e medida pela utilização da estatística de Durbin-Watson. Essa estatística mede a correlação entre cada resíduo e o resíduo para o período de tempo imediatamente antecedente àquele de interesse. Levine *et al.* (2000) definem a estatística de Durbin-Watson ( $D$ ) da seguinte forma:

$$D = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \quad (23)$$

onde  $e_i$  = resíduo no período de tempo  $i$ .

Para melhor compreender o que a estatística de Durbin-Watson está mensurando, faz-se necessário examinar a composição da estatística  $D$  na equação (23). O numerador representa a diferença ao quadrado entre dois resíduos sucessivos, somados desde a segunda observação até a  $n$ -ésima observação. O denominador representa a soma

dos quadrados dos resíduos. Quando resíduos sucessivos são positivamente autocorrelacionados, o valor de  $D$  irá se aproximar de 0. Se os resíduos não forem correlacionados, o valor de  $D$  estará próximo de 2.

### O modelo de Regressão Linear Múltipla (MRLM)

A análise de regressão múltipla é um estudo de como a variável dependente  $y$  está relacionada com duas ou mais variáveis independentes. Pode-se afirmar que o MRLM é uma extensão do MRLS. Anderson *et al.* (2002) descrevem o MRLM da seguinte forma:

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \dots + \beta_p \cdot x_p + \varepsilon \quad (24)$$

onde  $x_1, x_2, \dots, x_p$  são constantes,  $\beta_0, \beta_1, \beta_2, \dots, \beta_p$  são parâmetros denominados coeficientes de regressão parciais, e  $\varepsilon$  os resíduos.

Através da extensão do método dos mínimos quadrados empregado no MRLS, pode-se chegar à equação linear da regressão múltipla, isto é,

$$\hat{y} = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p \quad (25)$$

### O coeficiente de determinação múltipla

O coeficiente de determinação  $R^2$  pode ser interpretado como a proporção da variabilidade que pode ser estimada pela equação da regressão múltipla. Todavia, quando multiplicado por 100, o coeficiente de determinação múltipla pode ser interpretado como a porcentagem da variabilidade em  $y$  que pode ser explicada através da equação de regressão.

### O coeficiente de determinação múltipla ajustado

Anderson *et al.* (2002) observaram que muitos analistas preferem ajustar o  $R^2$  para o número de variáveis independentes, a fim de se evitar superestimar o impacto de adicionar uma variável independente na quantidade de variação explicada pela equação da regressão estimada.

A estatística  $R^2$  pode ser ajustada como:

$$R_a^2 = 1 - \left( \frac{SQReg}{SQT} \right) \cdot \left( \frac{n-1}{n-k-1} \right) \quad (26)$$

onde  $n$  representa o número de observações da amostra,  $k$  representa o número de variáveis independentes no modelo e  $R_a^2$ , o coeficiente de determinação múltipla ajustado (RAGSDALE, 2001).

### Testando a significância da relação entre a variável dependente e as variáveis explicativas

Como na regressão linear múltipla existe mais de uma

variável explicativa, as hipóteses nula e alternativa podem ser escritas da seguinte forma:

$H_0: \beta_1 = \beta_2 = \dots \beta_n = 0$  (não existe relação linear entre a variável dependente e as variáveis explicativas)

$H_1$ : Pelo menos um  $\beta_j \neq 0$  (pelo menos um coeficiente de regressão não é igual a 0)

Essa hipótese nula pode ser verificada através do teste  $F$ . Quando se testa a significância dos coeficientes de regressão, a medida do erro aleatório é chamada de variância do erro, de modo que o teste  $F$  é a fração da variância decorrente da regressão dividida pela variância do erro, conforme mostra a equação (27):

$$F = \frac{\text{QMReg}}{\text{QMRes}} \quad (27)$$

onde QMReg é igual ao Quadrado Médio devido à Regressão, e QMRes o Quadrado Médio devido à variação Residual.

A regra de decisão é rejeitar  $H_0$  no nível de significância  $\alpha$  se  $F > F_{(SP, n - P - 1)}$ , onde  $P$  é o número de variáveis explicativas no modelo de regressão, e  $F$  é a estatística de teste, a partir de uma distribuição  $F$  com  $n - P - 1$  graus de liberdade, conforme pode ser visto na Tabela 1.

**Teste para os parâmetros da população no MRLM**

Para testar a hipótese de que a inclinação da população  $\beta_j$  é igual a 0, utilizou-se a equação (20)

$$t = \frac{b_1}{S_{b_1}}$$

Por outro lado, esta expressão pode ser escrita de uma forma mais genérica para a regressão múltipla, como

$$t = \frac{b_k}{S_{b_k}} \quad (28)$$

onde  $S_{b_k}$  é o erro padrão do coeficiente de regressão  $b_k$ ,  $t$  corresponde à estatística  $t$  para um distribuição com  $n - P - 1$  graus de liberdade, sendo  $P$  o número de variáveis dependentes na equação de regressão múltipla.

**Multicolinearidade**

Para Ragsdale (2001), o termo multicolinearidade é utilizado para descrever uma situação onde as variáveis independentes num dado modelo de regressão estão correlacionadas entre si. Segundo Levine *et al.* (2000), nessas situações, variáveis colineares não fornecem novas informações, e torna-se difícil separar o efeito dessas variáveis na variável dependente ou na variável resposta. Nesses casos, os valores dos coeficientes de regressão para as variáveis correlacionadas podem flutuar drasticamente, dependendo se as variáveis estão ou não incluídas no modelo. Um método de mensuração da colinearidade utilizado refere-se ao Fator Inflacionário da Variância (FIV) para cada variável explicativa, o qual é definido na forma da equação (29):

$$\text{FIV}_j = \frac{1}{1 - R_j^2} \quad (29)$$

onde  $R_j^2$  é o coeficiente de determinação múltipla da variável explicativa  $X_j$  com todas as outras variáveis  $X$ , para  $j = 1, 2, 3, \dots, n$ .

Se existirem somente duas variáveis explicativas,  $R_j^2$  é somente o coeficiente de determinação entre  $X_1$  e  $X_2$ . Se, por exemplo, existissem três variáveis explicativas, então  $R_j^2$  seria o coeficiente de determinação múltipla de  $X_1$  com  $X_2$  e  $X_3$ .

Se um conjunto de variáveis explicativas não for correlacionado, então  $\text{FIV}_j$  será igual a 1. Se o conjunto for altamente correlacionado, então o  $\text{FIV}_j$  poderá até exceder a

**Tabela 1: ANOVA para testar a significância de um conjunto de coeficientes de regressão em um modelo de regressão múltipla.**

FONTE	GL	SOMA DOS QUADRADOS	QUADRADO DA MÉDIA ARITMÉTICA (VARIÂNCIA)	F
Regressão	P	SQReg	$\text{QMReg} = \frac{\text{SQReg}}{P}$	$F = \frac{\text{QMReg}}{\text{QMRes}}$
Erro	$n - P - 1$	SQR	$\text{QMRes} = \frac{\text{SQR}}{n - P - 1}$	
Total	$n - 1$	STQ		

Fonte: Levine *et al.*, página 591, 2000.

10. Marquardt *apud* Levine *et al.* (2000) sugere que se o  $FIV_j$  for maior que 10, então existe demasiada correlação entre a variável  $X_j$  e as outras variáveis explicativas. No entanto, outros pesquisadores sugerem um critério mais conservador, que empregaria alternativas para a regressão dos mínimos quadrados se o  $FIV_j$  exceder a 5.

## ANÁLISE DISCRIMINANTE

Para Ragsdale (2001) a análise discriminante é uma técnica estatística que usa a informação disponível em um grupo de variáveis independentes, com distribuição normal, para prever o valor de uma variável categórica dependente. Justamente pelo fato de a variável dependente ser categórica, ou discreta, a análise discriminante se diferencia da regressão.

Segundo Malhotra (2001), os objetivos da análise discriminante são: estabelecer funções discriminantes, ou combinações lineares das variáveis independentes ou prognosticadoras, que melhor discriminem entre as categóricas da variável dependente; verificar se existem diferenças significativas entre os grupos, em termos das variáveis prognosticadoras; determinar as variáveis preditoras que mais contribuem para as diferenças entre grupos; enquadrar ou classificar os casos a um dos grupos, com base nos valores das variáveis preditoras; e avaliar a precisão da classificação.

Para Ragsdale (2001), um modelo de análise discriminante envolve combinações lineares da seguinte forma:

$$\hat{Y}_i = b_0 + \sum_{i=1}^k b_i x_i \quad (30)$$

onde  $\hat{Y}_i$  é o escore discriminante,  $b$  o coeficiente de regressão ou peso discriminante e  $x$  a variável prognosticadora ou independente.

As estatísticas associadas à análise discriminante são:

- **Correlação Canônica** – mede a relação entre a função discriminante isolada e o conjunto de variáveis do grupo;
- **Centróide** – é a média dos valores dos escores discriminantes de um determinado grupo;
- **Matriz de Classificação** – é aquela que contém o total de casos classificados corretamente e mal classificados;
- **Coefficientes da função discriminante (não padronizados)** – são os multiplicadores de variáveis, quando as variáveis estão nas unidades de medida originais;
- **Escore Discriminante** – corresponde aos valores de  $\hat{Y}_i$ ;
- **Autovalor** – é a razão da soma de quadrados entre grupos para a mesma soma dentro de grupos. Grandes autovalores implicam funções superiores;
- **Teste F** – teste de significância;
- **$\lambda$  de Wilks ou estatística U** – é, para cada prognostica-

dor, a razão da soma e quadrados dentro dos grupos para a soma total de quadrados. Se  $\lambda \approx 1$ , então as médias dos grupos não diferem umas das outras; se  $\lambda \approx 0$ , elas diferem;

- **Médias de Grupos e desvios padrão de grupos** – são calculados para cada prognosticador e para cada grupo;
- **Coefficientes da função discriminante padronizada** – são usados como multiplicadores quando as variáveis foram padronizadas para média 0 e variância 1.
- **Correlações estruturais** – representam as correlações simples entre os prognosticadores e a função discriminante;
- **Matriz de correlação total** – se os casos são tratados como se proviessem de uma única amostra, calculando-se então as correlações, obtém-se uma matriz de correlação total.
- **Matriz estrutural** – Correlações combinadas dentro de grupos entre variáveis discriminantes e funções discriminantes canônicas (variáveis ordenadas por tamanhos da correlação dentro da função).

A análise discriminante pode envolver dois grupos ou mais. Para a realização da mesma, podem-se adotar os seguintes estágios:

- **Formulação do problema** – O primeiro passo da análise discriminante é formular o problema identificando os objetivos, a variável dependente e as variáveis independentes. A variável dependente deve consistir de duas ou mais categorias mutuamente excludentes e coletivamente exaustivas. Quando a variável dependente é escalonada em forma de intervalo ou razão, deve ser primeiramente convertida em categorias. As variáveis prognosticadoras devem ser selecionadas com base em um modelo teórico ou em uma pesquisa prévia. O próximo passo consiste em dividir a amostra em duas partes. Uma delas é chamada amostra de análise, que é utilizada para estimar a função discriminante. A outra é denominada amostra de validação, que é reservada para validar a função discriminante.
- **Estimação dos coeficientes da função discriminante** – Uma vez identificada a amostra de análise pode-se estimar os coeficientes da função discriminante. Existem dois métodos amplos de abordagem. O método direto consiste em estimar a função discriminante de modo que todos os prognosticadores sejam incluídos simultaneamente. Neste caso, todas as variáveis independentes são incluídas, independentemente de seu poder discriminatório. Este método é adequado quando, com base em uma pesquisa prévia ou em um modelo teórico, o pesquisador deseja que a discriminação esteja baseada em todos os prognosticadores. Uma abordagem alternativa é o método passo a passo. Na análise discriminante passo a passo, as

variáveis prognosticadoras são introduzidas seqüencialmente, com base em sua capacidade de discriminar entre os grupos. Este método é adequado quando o pesquisador deseja selecionar um subconjunto dos prognosticadores para inclusão na função discriminante.

- **Determinação da significância da função discriminante** – Pode-se testar estatisticamente a hipótese nula, de que as médias de todas as funções discriminantes em todos os grupos sejam iguais. Uma forma de realização desse teste é baseada no  $\lambda$  de Wilks. Testando-se simultaneamente várias funções (como é o caso da análise discriminante múltipla), a estatística  $\lambda$  de Wilks é o produto dos  $\lambda$  univariados de cada função. Estima-se o nível de significância com base em uma transformação qui-quadrado da estatística.
- **Interpretação e validação** – A interpretação dos pesos ou coeficientes da discriminante é similar à do caso da análise de regressão múltipla. O valor do coeficiente para um determinado prognosticador depende dos outros prognosticadores incluídos na função discriminante. Os sinais dos coeficientes são arbitrários, mas indicam que valores de variável resultam em grandes ou pequenos valores da função associando-os a grupos particulares. Dada a multicolinearidade nas variáveis prognosticadoras, não há medida não ambígua da importância relativa dos prognosticadores na discriminação dos grupos. Com esta precaução em mente, pode-se ter alguma idéia da importância relativa das variáveis, examinando a magnitude absoluta dos coeficientes padronizados da função discriminante. De modo geral, os prognosticadores com coeficientes padronizados relativamente grandes contribuem mais para o poder discriminatório da função, em comparação com prognosticadores com coeficientes menores.

## METODOLOGIA

Esta metodologia constou do desenvolvimento das seguintes etapas:

a) Avaliação das condições climáticas em ambientes de trabalho. Utilizaram-se equipamentos do Laboratório de Análise do Trabalho (LAT) do Departamento de Engenharia de Produção (DEP) da Universidade Federal da Paraíba (UFPB), que atendem às exigências da Norma ISO-DIS 7726/1996 (Ergonomics of The Thermal Environment – Instruments for Measuring Physical Quantities). Questionários foram distribuídos a todas as pessoas participantes da pesquisa, para registrarem as respectivas opiniões sobre sensação térmica, tipo de roupa utilizada e dados pessoais, conforme a Norma ISO 10551/1995 (Subjective Judgment Scales).

b) Verificação da normalidade das amostras coletadas através das medidas de BOX-COX. Trata-se de um teste de

normalidade de Bera-Jarque, o qual é uma consequência do estudo feito por Shenton e Bowman em 1977, que é construído com as expressões para assimetria e curtose, de acordo com a equação (31):

$$N \left[ \frac{\sqrt{b_1^2}}{6} + \frac{(b_2 - 3)^2}{24} \right] \approx \chi^2 \quad (31)$$

onde:  $\sqrt{b_1^2}$  = Assimetria;  $b_2$  = Curtose; e  $\chi^2$  = Qui-quadrado.

Quando  $b_1$  e  $b_2$  são grandes, a expressão acima também é, e, por sua vez, maior que o Qui-quadrado ( $\chi^2$ ) tabelado. Ou seja, se o teste aplicado às variáveis coletadas ultrapassar o valor de 5,91, que representa 2 graus de liberdade e uma frequência de 0,95 na tabela do Qui-quadrado, a amostra relativa à variável testada não obedecerá a uma curva normal.

c) Construção de modelo de regressão linear. Segundo Tabachnick e Fidel (1989), Levine *et al.* (2000), Charnet *et al.* (1999) e Ragsdale (2001), o modelo geral de uma equação linear pode ser expresso como uma variável dependente (VD) em função de um conjunto de variáveis independentes (VI), ou seja, pode-se expressar o valor esperado da variável dependente (VD) como função de várias variáveis regressoras (VI). Desta forma utilizou-se o modelo apresentado na equação (32), a saber:

$$y_p = \beta_0 + \sum_{i=1}^n \beta_i x_i + \varepsilon \quad (32)$$

onde:

$y_p$  = Variável dependente ( $S$  – sensação térmica) predita pela variável de conforto térmico;

$\beta_0$  = Constante de regressão ou intercepto;

$\beta_i$  = Coeficiente parcial ou parâmetro da regressão para a variável  $x_i$ ;

$x_i$  = Variáveis térmicas;

$\varepsilon$  = Erro ou resíduo devido ao modelo de regressão (estatístico e não-determinístico).

d) Avaliação da consistência da equação e de seus parâmetros através dos testes  $t$  e  $F$ .

e) Determinação da temperatura de bulbo seco e da velocidade relativa do ar de conforto térmico para os ambientes estudados, segundo a percepção das pessoas.

f) Realização da análise discriminante, com o intuito de investigar sobre as possíveis características existentes entre as variáveis térmicas e as sensações declaradas pelos trabalhadores, segundo esta classificação: conforto (0), quente (1) e muito quente (2).

## RESULTADOS E DISCUSSÕES

Nos ambientes de trabalho estudados realizam-se atividades de operações bancárias, ou seja, atividade sedentária e vestimentas leves. Durante três dias foram coletados 5.496 dados inerentes às variáveis climáticas  $I_{cl}$  (resistência térmica das vestes),  $M$  (metabolismo),  $t_{bs}$  (temperatura do ar ou de bulbo seco),  $t_{rm}$  (temperatura radiante média),  $t_{bu}$  (temperatura de bulbo úmido),  $V_{ar}$  (velocidade relativa do ar),  $UR$  (umidade relativa do ar em percentual) e  $t_{op}$  (temperatura operativa); e  $S$  (sensação térmica). Importa ressaltar que as unidades atribuídas às variáveis  $I_{cl}$  e  $V_{ar}$  são, respectivamente, [ $\text{clo} = 0,155\text{m}^2\cdot\text{C}/\text{W}$ ] e [ $\text{m}/\text{s}$ ]. A temperatura de bulbo seco e a velocidade relativa do ar, na situação de conforto térmico, foram determinadas a partir da coleta destes dados e levaram-se em consideração os votos das sensações térmicas fornecidos pelas pessoas nos diversos horários.

Dando seqüência às etapas propostas, verificou-se através das medidas de BOX-COX a normalidade das variáveis acima, de acordo com a equação (31), conforme pode ser visto na Tabela 2 a seguir.

Verificada a normalidade das variáveis,  $BJ < 5,91$ , aplicaram-se a Regressão Linear Simples e Múltipla e a Análise Discriminante.

### Regressão Linear Simples

Depois de realizados os devidos testes, elaborou-se um

modelo de Regressão Simples, contendo a sensação térmica ( $S$ ) como variável dependente e a temperatura de bulbo seco ( $t_{bs}$ ) como variável independente.

A Tabela 3 apresenta os resultados da Regressão Linear Simples onde  $S$  é função da variável independente  $t_{bs}$ .

Os testes  $F$  e  $t$  mostram a consistência do modelo para  $\alpha = 0,005$  e  $\alpha = 0,10$  ( $p$ -valor = 0,00000), respectivamente. A Figura 1 ilustra a reta de regressão obtida e os valores observados.

Observa-se, também, que o valor de  $R^2$  é próximo de 1, donde conclui-se que o modelo é consideravelmente explicado pela variável preditora.

Através da Figura 2, observa-se o comportamento dos resíduos padronizados ao longo das medições. *A priori*, não se percebe nenhuma espécie de padrão dos resíduos no gráfico abaixo.

A estatística de Durbin-Watson corrobora a afirmação anterior mostrando que não há indícios de autocorrelação entre os resíduos, já que o valor desta é 1,528, ou seja, maior que o valor crítico  $d'_s$  (para  $p = 1$ ,  $n = 30$ ) tabelado, que corresponde a 1,49. Logo, a partir dos valores de B constantes na Tabela 3, o MRLS pode ser expresso como

$$S = -10,5531 + 0,4531 \cdot t_{bs} \quad (33)$$

A partir deste modelo, é possível prever que a temperatura de bulbo seco na situação de conforto térmico ( $S = 0$ ) é de 23,29 °C.

Tabela 2: Valores experimentais dos coeficientes (Distorção, Curtose e Bera-Jarque) relativos a cada uma das variáveis envolvidas na pesquisa.

VARIÁVEIS COEFICIENTE	$I_{cl}$	$t_{bs}$	$t_{bu}$	$t_{rm}$	$V_{ar}$	$UR$	$t_{op}$	$S$
Coef. Distorção	-0,79083	1,015	2,408	0,7473	2,796	3,0206	0,9448	0,02
Coef. Curtose	1,366008	-0,4	5,357	0,062	7,964	8,3185	-0,113	-0,72
Bera-Jarque (BJ)	0,209554	0,176	1,597	0,1247	3,109	3,3867	0,158	0,02

Tabela 3: Resultados da Regressão Linear Simples entre  $S$  e  $t_{bs}$ .

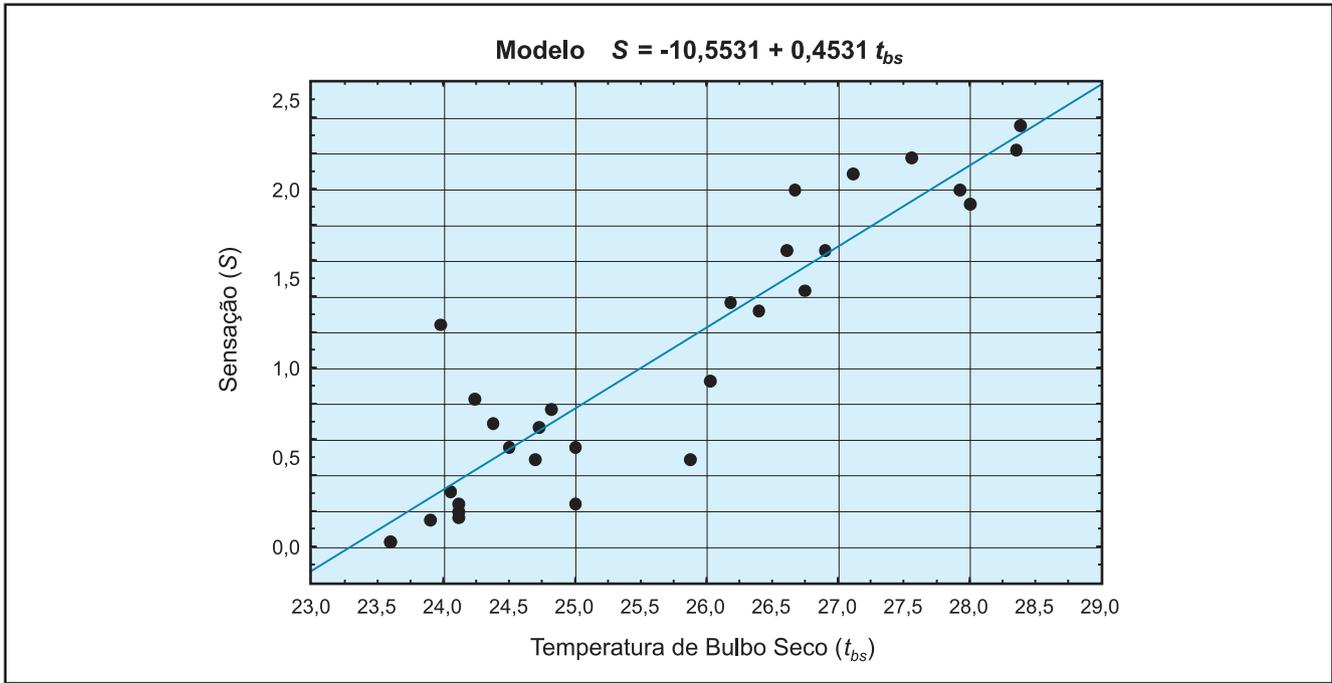
$R$	$R^2$	$R^2(\text{AJUSTADO})$	ERRO PADRÃO		$F(1,28)$	$P$ -VALOR
0,9187	0,8440	0,8384	0,3059		151,4604	0,0000
	Beta	Erro Padrão de Beta	B	Erro Padrão de B	$t(28)$	$p$ -valor
Intercepto			-10,5531	0,943150	-11,1892	0,00000
$t_{bs}$	0,918682	0,014648	0,4531	0,036815	12,3069	0,00000

**Regressão Linear Múltipla**

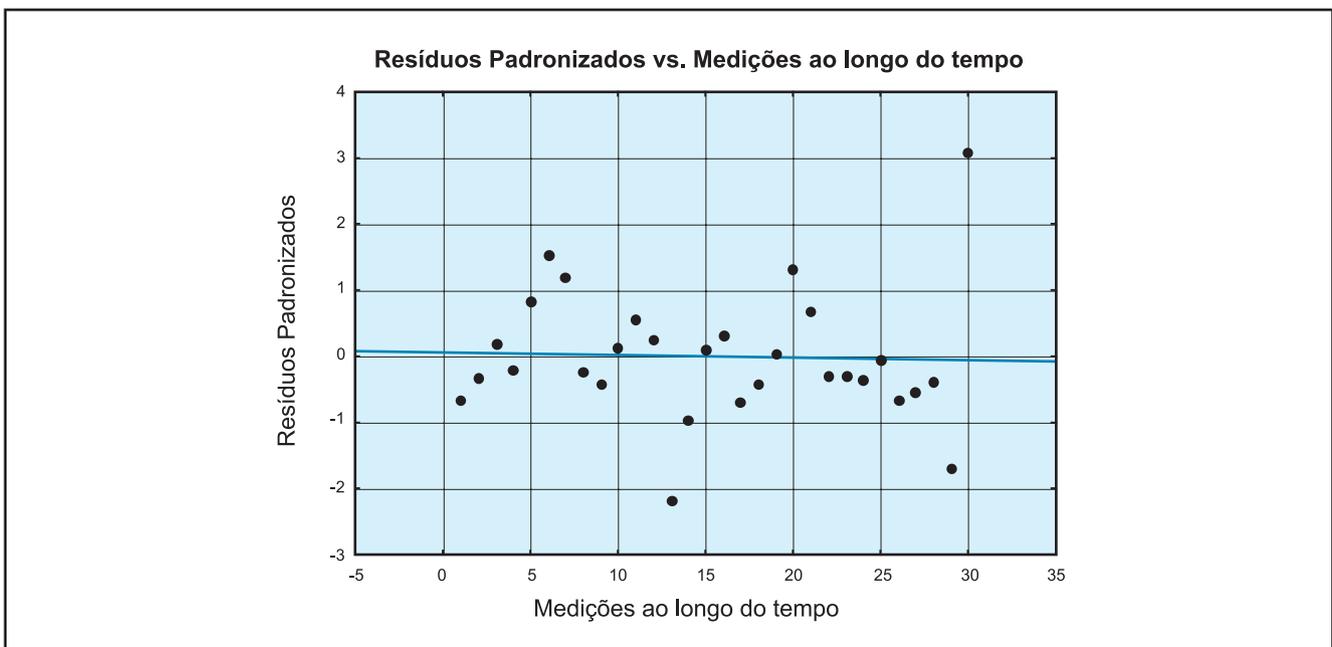
Baseando-se nos resultados e discussões anteriores, construiu-se apenas um modelo de regressão múltipla. Este teve a sensação térmica ( $S$ ) como variável dependente e as

variáveis independentes temperatura de bulbo seco ( $t_{bs}$ ) e velocidade relativa do ar ( $V_{ar}$ ). A variável  $V_{ar}$  foi selecionada devido a sua importância na sensação de conforto térmico do trabalhador.

**Figura 1: Retas de regressão entre as temperaturas de bulbo seco e as respectivas sensações relatadas pelas pessoas.**



**Figura 2: Gráfico dos resíduos padronizados em relação às medições feitas ao longo do tempo (MRLS).**



Antes de construir o MRLM verificou-se a correlação entre as variáveis independentes. O valor encontrado foi de 0,275961, ou seja, um valor não muito significativo. Concluiu-se, então, que não há existência de multicolinearidade entre essas variáveis.

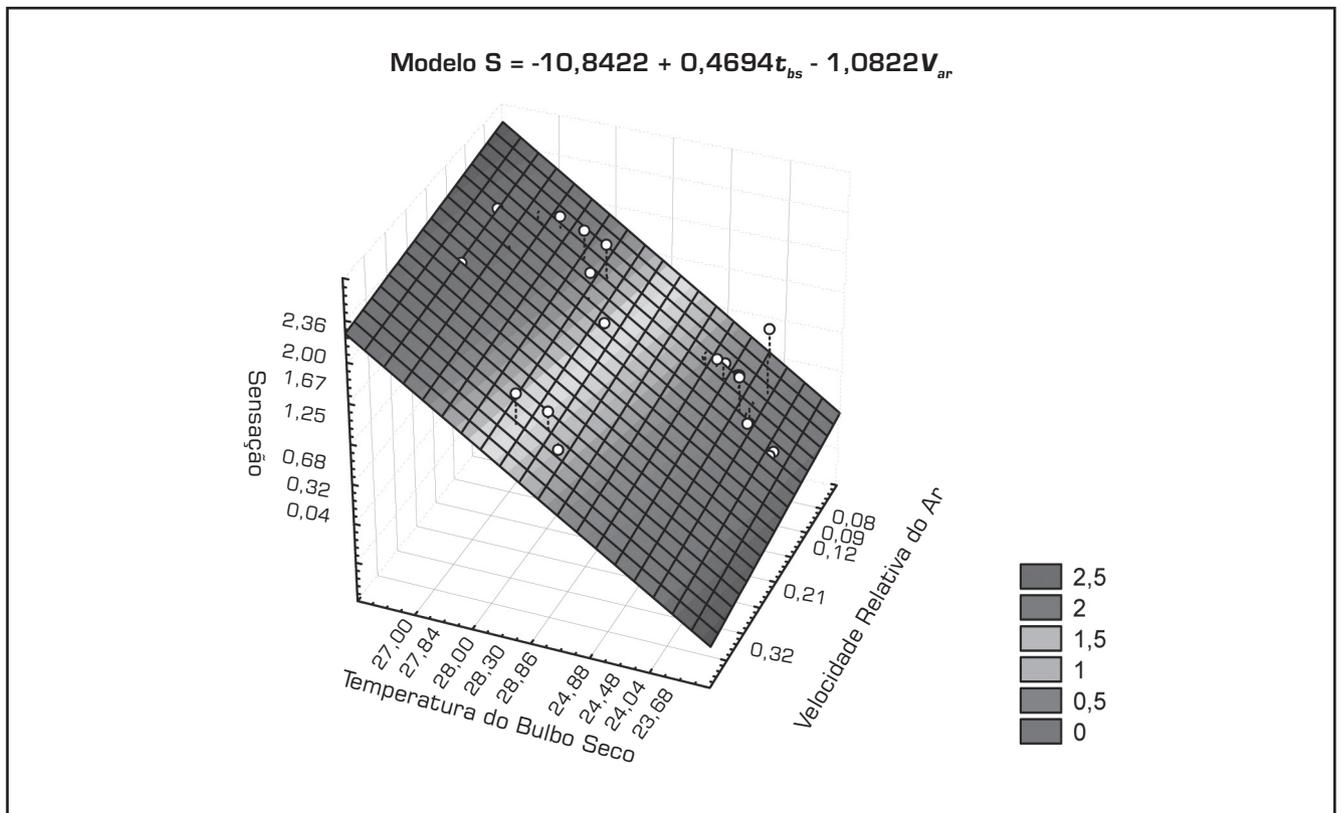
Em seguida, construiu-se o MRLM. Os resultados da Regressão Linear Múltipla encontram-se na Tabela 4.

O teste *F* indica que o modelo é significativamente explicado pelo conjunto de variáveis independentes ao nível de  $\alpha = 0,05$  (*p*-valor = 0,00000). No entanto, o teste *t* mostra

Tabela 4: Resultados da Regressão Linear Múltipla entre  $S(VD)$ ,  $t_{bs}(VI)$  e  $V_{ar}(VI)$ .

R	R <sup>2</sup>	R <sup>2</sup> (AJUSTADO)	ERRO PADRÃO		F(2,27)	P-VALOR
0,92611	0,85767	0,84713	0,29753		81,35331	0,0000
	Beta	Erro Padrão de Beta	B	Erro Padrão de B	t(27)	p-valor
Intercepto			-10,8422	0,934691	-11,5997	0,000000
$t_{bs}$	0,952285	0,075537	0,4696	0,037253	12,6069	0,000000
$V_{ar}$	-0,12177	0,075537	-1,0822	0,671321	-1,6120	0,118584

Figura 3: Superfície de regressão entre as temperaturas de bulbo seco, velocidades relativas do ar e as respectivas sensações relatadas pelas pessoas.



que apenas o intercepto e a variável  $t_{bs}$  são significativos ao nível de  $\alpha = 0,10$ . Mas como  $p\text{-valor} > \alpha$  para variável  $V_{ar}$  em 1,8584% e o  $R^2$  (ajustado) apresentado na Tabela 4 é 0,84713, superior ao valor de 0,8384 constante na Tabela 3, preferiu-se incluir a variável  $V_{ar}$  no modelo, proporcionando assim uma correlação múltipla maior entre  $S$  e o conjunto de variáveis  $t_{bs}$  e  $V_{ar}$ . É possível ver na Figura 3 que a sensação térmica ( $S$ ) do indivíduo tende a situação de conforto ( $S = 0$ ) quando as variáveis  $t_{bs}$  e  $V_{ar}$  se aproximam para valores ínfimos.

Pelo gráfico exposto na Figura 4, não é possível identificar claramente evidências indicando a existência de autocorrelação entre os resíduos. O valor da estatística de Durbin-Watson encontrado foi 1,679740, isto é, superior ao valor crítico tabelado  $d_s$  (para  $p = 2$ ,  $n = 30$ ), elucidando assim a inexistência de autocorrelação entre os resíduos, o que ratifica a permanência da variável  $V_{ar}$  na equação (34).

Verificada a significância e consistência do modelo e a partir dos valores de B da Tabela 4 tem-se que o mesmo é dado pela equação

$$S = -10,8422 + 0,4694 \cdot t_{bs} - 1,0822 \cdot V_{ar} \quad (34)$$

Para averiguar a situação de conforto térmico ( $S = 0$ )

através da relação encontrada na equação (34) faz-se necessário fixar o valor de uma das variáveis independentes, a fim de se determinar o valor da outra variável preditora. Sendo assim, adotando-se o valor de  $t_{bs}$  na equação (34) como sendo 23,29°C, obtido na equação (33) verifica-se que para situação de conforto térmico o valor da variável  $V_{ar}$  é de 0,0823 m/s.

### Análise Discriminante

Como visto anteriormente, sabe-se que um modelo de Análise Discriminante possui uma variável dependente categórica, isto é, discreta. Novamente a variável dependente em questão é a Sensação ( $S$ ), no entanto esta foi categorizada em três diferentes grupos<sup>2</sup>, a saber:

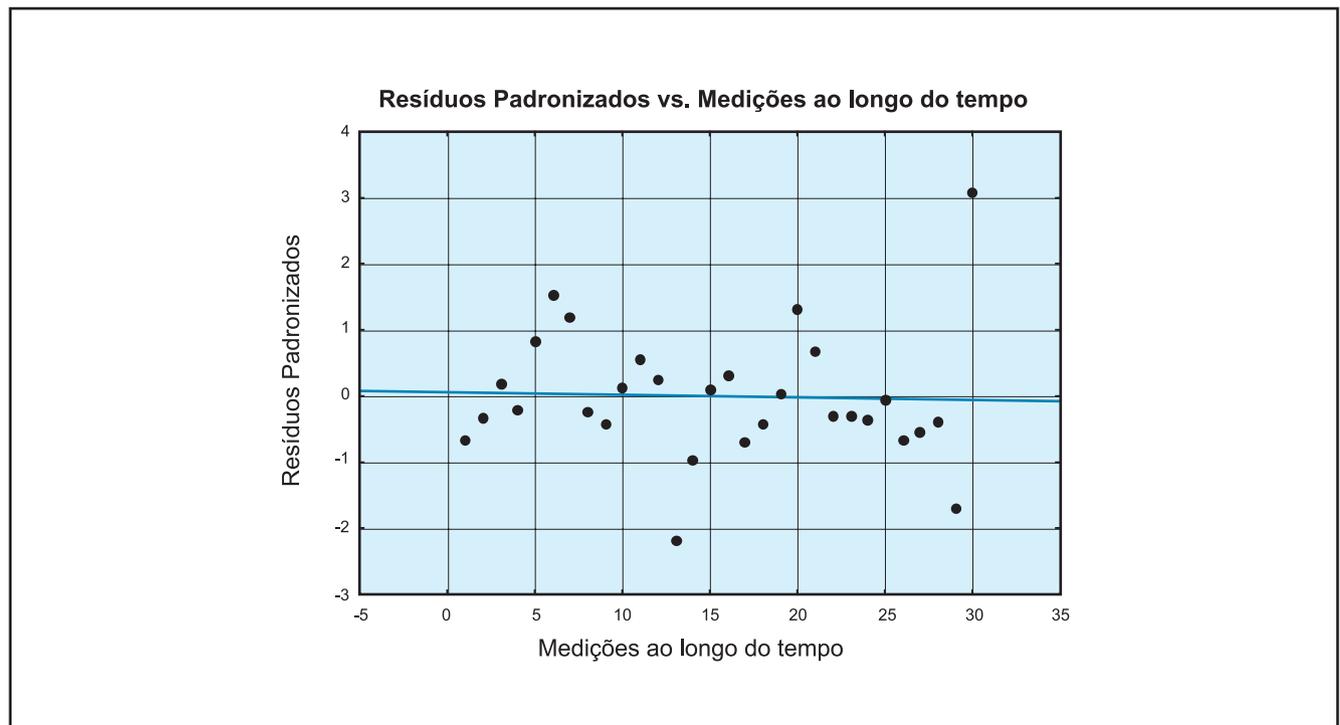
0 – Conforto Térmico; 1 – Quente; 2 – Muito Quente.

As variáveis independentes são: Temperatura de Bulbo Seco ( $t_{bs}$ ) e Velocidade Relativa do Ar ( $V_{ar}$ ).

Os resultados a serem apresentados a seguir foram obtidos através dos Softwares STATISTICA e SPSS. Importa ressaltar que ambos apresentaram pequenas diferenças em seus resultados, todavia estas não comprometem a validação dos resultados.

Os resultados apresentam uma análise discriminante de

Figura 4: Gráfico dos resíduos padronizados em relação às medições feitas ao longo do tempo (MRLM).



### Médias dos Grupos

Grupo	$t_{bs}$	$V_{ar}$
0	24,50324	0,11294
1	26,39000	0,17143
2	27,65500	0,10250
TOTAL	25,57383	0,12450

### Desvio Padrão dos Grupos

Grupo	$t_{bs}$	$V_{ar}$
0 N = 17	0,690156	0,079903
1 N = 7	1,219686	0,111718
2 N = 6	0,687306	0,054749
TOTAL N = 30	1,542996	0,085625

### Matriz de Correlação Combinada Dentro de Grupos

	$t_{bs}$	$V_{ar}$
$t_{bs}$	1,000	0,44
$V_{ar}$	0,447	1,000

### $\lambda$ de Wilks e Razão F Univariada Com 2 e 27 Graus de Liberdade

Variável	$\lambda$ de Wilks	F	Significância
$t_{bs}$	0,274	35,758	0,000
$V_{ar}$	0,904	1,434	0,256

### Funções Discriminantes Canônicas

Função	Autovalor	Porcentagem	Porcentagem de variação	Correlação Canônica
1	3,213	96,8	96,8	0,873
2	0,105	3,2	100,00	0,308
"After Function" (Função após)	$\lambda$ de Wilks	$Qui^2$	Graus de liberdade	p-valor
0	0,215	40,749	4	0,000
1	0,905	2,635	1	0,105

**Coefficientes da Função Discriminante Canônica Padronizada**

	<b>Função 1</b>	<b>Função 2</b>
$t_{bs}$	1,118	-0,023
$V_{ar}$	-0,479	1,010

**Matriz Estrutural**

	<b>Função 1</b>	<b>Função 2</b>
$t_{bs}$	0,905(*)	0,426
$V_{ar}$	-0,479	1,000(*)

(\*) Maior correlação absoluta entre cada variável e qualquer função discriminante

**Coefficientes Não-Padronizados da Função Discriminante Canônica**

	<b>Função 1</b>	<b>Função 2</b>
<b>Intercepto</b>	-33,465	-0,769
$t_{bs}$	1,336	-0,028
$V_{ar}$	-5,689	11,981

**Centróides dos Grupos**

<b>Grupo</b>	<b>Função 1</b>	<b>Função 2</b>
0	-1,365	-0,108
1	0,824	0,539
2	2,906	-0,322

**Coefficientes das Funções de Classificação**

	<b>Sensação (S)</b>		
	0 p = 0,5667	1 p = 0,2333	2 p = 0,2000
<b>Intercepto</b>	-515,983	-590,151	663,122
$t_{bs}$	42,874	45,780	48,587
$V_{ar}$	-174,698	-179,392	-201,561

**Matriz de Classificação**

Grupo	0	1	2	Porcentagem de classificações corretas
0	17	0	0	100,0%
1	1	4	2	57,1%
2	0	0	6	100,0%
Total	18	4	8	<b>90,0%</b>

**Matriz de Classificação para Validação Cruzada**

Grupo	0	1	2	Porcentagem de classificações corretas
0	15	2	0	88,2%
1	2	3	2	42,9%
2	0	1	5	83,3%
Total	17	6	6	<b>76,7%</b>

três grupos. Aparentemente, os três grupos são mais separados em termos da variável  $t_{bs}$  do que da outra variável ( $V_{ar}$ ), conforme observada nas médias e desvio padrão dos grupos. A matriz de correlação combinada dentro dos grupos indica fraca correlação entre os prognosticadores. É improvável que a multicolinearidade venha a constituir um problema. A significância  $F$  univariada indica que, quando os prognosticadores são considerados individualmente, apenas a variável  $t_{bs}$  distingue significativamente a sensação térmica percebida pelos indivíduos.

Como há três grupos, estima-se apenas duas funções discriminantes (1 e 2). Os autovalores associados a estas funções são 3,213 e 0,105, com correlação de 0,873 e 0,308, respectivamente. Os quadrados destas correlações  $(0,873)^2 \cong 0,76$  e  $(0,308)^2 \cong 0,10$  indicam que 76% e 10% das variâncias na variável dependente ( $S$ ) são explicados pelos modelos 1 e 2 respectivamente.

Observando os resultados das funções discriminantes canônicas, o valor 0 localizado abaixo da “After Function” mostra que as duas funções 1 e 2 conjuntamente tornam-se uma função discriminante significativa ( $\alpha = 0,05$ ), haja vista que o  $\lambda$  de Wilks associado a esta função é 0,215, que se transforma em um qui-quadrado 40,749 com 4 graus de liberdade, e  $p$ -valor = 0,000. Desta maneira, pode-se afirmar que a hipótese nula foi rejeitada, o que confirma a discriminação significativa. Porém, ao se remover a primeira função, constata-se que a segunda função não exerce influência

significativa para a diferenciação dos grupos ( $p$ -valor = 0,105).

Os coeficientes padronizados da função discriminante canônica ilustram um maior coeficiente para a  $t_{bs}$  na função 1, ao passo que na função 2 o coeficiente da  $V_{ar}$  possui um valor maior. Através da matriz estrutural também é plausível chegar a uma conclusão semelhante, ou seja, a variável  $t_{bs}$  possui peso maior na função 1, e a variável  $V_{ar}$ , maior peso na função 2.

Os resultados observados na matriz de classificação demonstram que 90% das classificações entre os grupos 1 e 2 foram realizadas de maneira correta, enquanto que na matriz de classificação para validação cruzada, as classificações se apresentaram corretas em 76,7% dos casos.

De acordo com o diagrama apresentado na Figura 5, nota-se que os grupos de percepção 0 (conforto) e 2 (muito quente) apresentam pontos mais aglomerados, ao contrário dos pontos do grupo 1 (quente) que se encontram mais dispersos. Isso explica o melhor desempenho no tocante a classificação dos grupos 0 e 2 em relação ao grupo 1.

**CONCLUSÃO**

A partir de um estudo baseado em autores da área métodos quantitativos, pôde-se observar e prever como uma ou mais variáveis podem influenciar na variabilidade de outra variável. Determinou-se, assim, a temperatura de bulbo seco e a

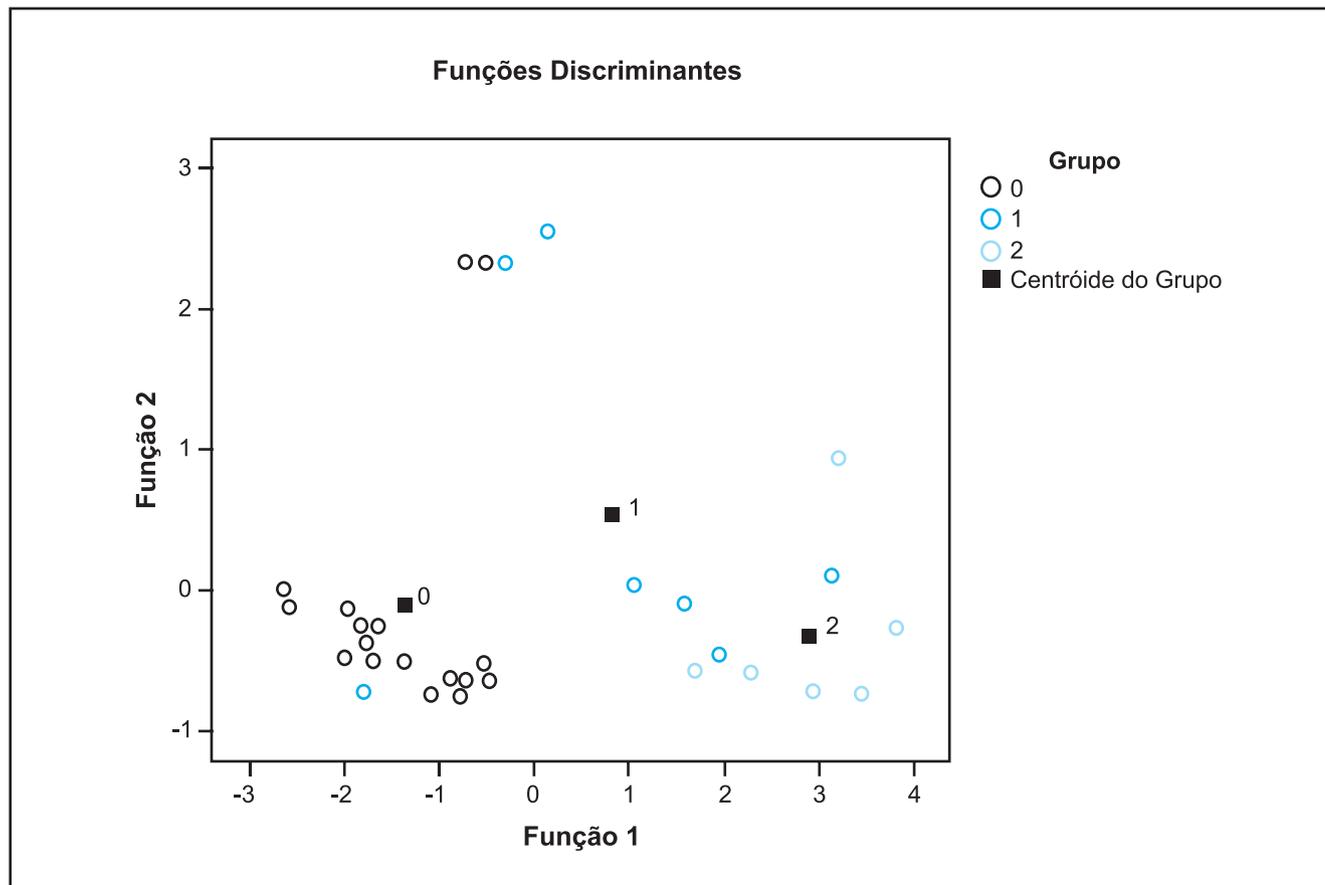
velocidade relativa do ar ideais para um ambiente de trabalho, baseando-se na sensação térmica de cada trabalhador, ou seja, construíram-se modelos de Regressão Linear Simples e Múltipla para determinar que variável térmica poderia estar em conformidade com a situação de conforto térmico ( $S = 0$ ) das pessoas que realizam atividades de bancos em ambientes climatizados, ou seja, atividade e vestimentas leves e umidade relativa no entorno de 50%. A temperatura de bulbo seco  $t_{bs} = 23,79^{\circ}\text{C}$  é a mais representativa e a ideal para estes ambientes.

As particularidades entre as sensações térmicas declaradas pelas pessoas foram constatadas e, através da Análise Discriminante, construíram-se funções de classificação segundo as preferências pelo conforto (0), quente (1) e pelo muito quente (2), as quais ratificam a importância da variável  $t_{bs}$  na discriminação destas preferências, demonstrando uma maior consistência do modelo  $S = -10,5531 + 0,4531 \cdot t_{bs}$ . Para facilitar a classificação de grupos discriminantes vinculados à sensação térmica, a partir de novas coletas de dados termoambientais, apresenta-se, anexo, um *script* desenvolvido no ambiente *MATLAB*.

Porém, vale ressaltar que muitos ambientes de trabalho, como os do setor bancário, estão localizados em regiões onde a temperatura externa do ar, em alguns períodos do ano, ultrapassa as temperaturas de projeto recomendadas pelas normas específicas, implicando aumento da temperatura interna, principalmente quando o teto não é bem isolado. Além disso, é comum verificar-se aumento eventual no número de pessoas e de equipamentos eletrônicos não previsto no projeto. Por outro lado, aberturas eventuais de portas e/ou janelas e, principalmente, os dispositivos de insuflamento do ar podem gerar velocidades tais que causem desconforto térmico.

De modo geral, o controle da carga térmica é de grande importância para a saúde e segurança do trabalhador, com repercussões na produtividade. O excesso ou o *déficit* de calor pode causar desconforto às pessoas com implicação na *performance* em atividades que exigem certa atenção e concentração. Assim, é importante ratificar que o modelo e as funções de classificação supracitadas poderão contribuir para o controle do conforto térmico interno, levando em consideração a percepção térmica das pessoas presentes no ambiente de trabalho.

Figura 5: Diagrama de dispersão de todos os grupos da função 1 sobre a função 2.



Artigo recebido em 21/12/2005

Aprovado para publicação em 21/08/2006

## ■ Notas

1. O erro padrão da estimativa é a medida de variabilidade (desvio padrão) em torno da linha de regressão.

2. O critério de discriminação da variável dependente adotado baseou-se nas características dos dados coletados.

## ■ Referências

ANDERSON, D. R.; SWEENEY, D. J.; & WILLIAMS, T. *Statistics for business and economics*. 3 ed, South-Western, 2002

ANSI/ASHRAE 55. *Thermal environmental conditions for human occupancy*. Atlanta, GA 30329, 1992.

BEDALE, E. M. The effects of posture and rest in muscular work. UK: *Industrial Fatigue Research*, 1924.

CHARNET, R.; BONVINO, H; FREIRE C. A. L; CHARNET, E. M. R. *Análise de modelos de regressão linear com aplicações*. Campinas, SP: Unicamp, 1999.

FANGER, P. O. *Thermal comfort – analysis and applications in environmental engineering*. United States: McGraw-Hill Book Company, 1970.

ISO 8996. Ergonomics – determination of metabolic heat production. *International Organization for Standardization*, Genebra, 1990.

ISO 9920. Ergonomics of the thermal environment – estimation of the thermal insulation and evaporative resistance of a clothing ensemble. *International Organization for Standardization*, Genebra, 1995.

ISO 10551. Ergonomics of the thermal environment – assessment of the influence of the thermal environment using subjective judgement scales. *International Organization for Standardization*, Genebra, 1995.

ISO/DIS 7726. Ergonomics of the thermal environment – Instruments for measuring physical quantities. *International Organization for Standardization*, Genebra, 1996.

ISO 7243. Hot environments – estimation of the heat stress on working man, based on the WBGT-index (wet bulb globe temperature). *International Organization for Standardization*, Genebra, 1989.

ISO 7730. Moderate Thermal Environments – Determination of the PMV and PPD indices and specification of the conditions for thermal comfort. *International Organization for Standardization*, Genebra, 1994.

LEVINE, D.M.; BERENSON, M. L.; STEPHAN, D. *Estatística; Teoria e aplicações*. Rio de Janeiro, LTC, 2000.

MAGALHÃES, M; LIMA, A. C. P. *Noções de probabilidade e estatística*. 5. ed., São Paulo: Edusp, 2002.

MALHOTRA, Naresh K. *Pesquisa de marketing – uma orientação aplicada*. Porto Alegre: Bookman, 3. ed., 2001.

NELSON, T.M.; NILSSON, T.H.; HOPKINS, G. W. Thermal comfort: advantages and deviations. Atlanta: *ASHRAE Transactions*, v. 93, n. 1, p. 1039-1054, 1987.

RAGSDALE, C.T. *Spreadsheet Modeling and Decision Analysis*, 3 ed, South-Western College Publishing, Cincinnati, Ohio, 2001.

RUAS, A. C. Avaliação de conforto térmico – contribuição à aplicação prática das normas internacionais. *Dissertação de Mestrado* – Faculdade de Engenharia Civil, Universidade Estadual de Campinas, Campinas, SP, 1999.

SHENTON, L.R.; BOWMAN, K.O. A bivariate model for the distribution of  $\sqrt{b_1}$  and  $b_2$ . *Journal of the American Statistical Association*, v. 72, p. 206-211, 1977.

TRIOLA, Mario F. *Introdução à Estatística*. 7. ed., Rio de Janeiro: LTC, 1999.

XAVIER, A. A. de Paula. Condições de conforto térmico para estudantes de 2ª Grau na região de Florianópolis. *Dissertação de mestrado*. CPGEC da UFSC, Florianópolis, 1999.

## ■ Sobre os autores

### Anand Subramanian

Universidade Federal da Paraíba (UFPB)

Engenheiro de Produção Mecânica. Mestrando em Engenharia de Produção

End.: Departamento de Engenharia de Produção (DEP) – Centro de Tecnologia – Campus I – Bloco G – Cidade Universitária, Castelo Branco, CEP: 58051-970 – João Pessoa, PB – Brasil – Caixa Postal: 5045.

Tel.: (83) 3216-7685, Fax: (83) 3216-7549

E-mail: anandsubraman@hotmail.com

### Antonio Souto Coutinho, Dr.

Universidade Federal da Paraíba (UFPB)

Professor Adjunto IV

End.: Departamento de Engenharia de Produção (DEP) – Centro de Tecnologia – Campus I – Bloco G – Cidade Universitária, Castelo Branco, CEP: 58051-970 – João Pessoa, PB – Brasil – Caixa Postal: 5045.

Tel.: (83) 3216-7685, Fax: (83) 3216-7549

E-mail: coutinho@ct.ufpb.br

### Luiz Bueno da Silva, Dr.

Universidade Federal da Paraíba (UFPB)

Professor Adjunto III

End.: Departamento de Engenharia de Produção (DEP) – Centro de Tecnologia – Campus I – Bloco G – Cidade Universitária, Castelo Branco, CEP: 58051-970 – João Pessoa, PB – Brasil – Caixa Postal: 5045.

Tel.: (83) 3216-7685, Fax: (83) 3216-7549

E-mail: bueno@ct.ufpb.br

**SCRIPT DESENVOLVIDO NO AMBIENTE *MATLAB* PARA CLASSIFICAÇÕES DE GRUPOS DISCRIMINANTES A PARTIR DE DADOS PROVENIENTES DE NOVAS AMOSTRAS**

A seguir segue um exemplo de aplicação do *script* para uma amostra de tamanho  $N = 6$ . Vale salientar que os dados de  $t_{bs}$  e  $V_{ar}$  foram aproveitados dos dados existentes da amostra utilizada neste estudo.

```
% Tamanho da Amostra
N = 6;
% Variaveis Independentes
Tbs = [27.10 26.59 24.68 23.58 24.04 24.11];
Var = [0.08 0.34 0.07 0.12 0.11 0.09];
% Funções de classificação
F0 = 42.874*Tbs - 174.698*Var - 515.983;
F1 = 45.780*Tbs - 179.392*Var - 590.151;
F2 = 48.587*Tbs - 201.561*Var - 663.122;
S = [];
for i = 1:N
F = [F0(i) F1(i) F2(i)];
C = max(F);
if F0(i) == C
    S(i) = 0;
elseif F1(i) == C
    S(i) = 1;
else
    S(i) = 2;
end
end
% Variavel Dependente
Sensacao = S
    Depois de executado o script, se obteve o seguinte resultado:

    Sensação térmica = 2 1 0 0 0 0
```