

**UNIVERSIDADE FEDERAL DO PARANÁ
SETOR DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA**

**BRUNO GERONYMO - GRR20159232
HERMANN MOGIZ DELGADO - GRR20159211
MARIA HELENA OLIVEIRA - GRR20159213
VINICIUS CESAR PEDROSO – GRR20159897**

ESTATURA BASEADA NO COMPRIMENTO DA MÃO E DO PÉ

CE225 – MODELOS LINEARES GENERALIZADOS

**CURITIBA
AGOSTO DE 2017**

Resumo

A altura de um indivíduo pode ser medida com muita precisão utilizando-se de instrumentos adequados. Buscando explorar as proporções de membros do corpo humano, foi realizado um estudo envolvendo 3 fatores que podem estar associados à estatura de uma pessoa, são eles o comprimento em milímetros da mão e do pé e também o gênero (Masculino e Feminino). Para tanto foi utilizado uma análise de dois modelos distintos de regressão linear múltipla, contemplando a variável resposta (estatura) em função das variáveis comprimento da mão e do pé, uma vez que a variável gênero foi utilizada para dividir a modelagem. Com a inclusão de termos quadráticos, o modelo linear múltiplo apresentou bom ajuste, determinando que há forte relação entre as variáveis em estudo.

Palavras-chave: modelo de regressão linear; altura; termos quadráticos.

Lista de Figuras

| | |
|--------------------------------------------------------------------------------------------------------|----|
| Figura 1: Gráfico de dispersão entre as variáveis <i>comprimento da mão</i> e <i>altura (mm)</i> | 6 |
| Figura 2: Gráfico de dispersão entre as variáveis <i>comprimento do pé</i> e <i>altura (mm)</i> | 6 |
| Figura 3: Gráfico de dispersão entre as variáveis <i>gênero</i> e <i>altura (mm)</i> | 7 |
| Figura 4: Gráfico de densidade da variável <i>altura (mm)</i> | 7 |
| Figura 5: Gráfico de densidade da variável <i>comprimento da mão (mm)</i> | 8 |
| Figura 6: Gráfico de densidade da variável <i>comprimento do pé (mm)</i> | 8 |
| Figura 7: Gráfico de resíduos | 9 |
| Figura 8: Gráfico de resíduos | 11 |

Sumário

| | |
|---------------------------------------------------------|----|
| 1 INTRODUÇÃO | 5 |
| 2 MATERIAL E MÉTODOS | 5 |
| 2.1 Material | 5 |
| 2.2 Métodos | 5 |
| 3 RESULTADOS E DISCUSSÃO..... | 6 |
| 3.1 Análise Descritiva e Exploratória..... | 6 |
| 3.2 Ajuste do Modelo de Regressão | 8 |
| 3.2.1 Modelo de Regressão para o gênero feminino | 8 |
| 3.2.2 Modelo de Regressão para o gênero masculino | 10 |
| 4 CONSIDERAÇÕES FINAIS | 12 |

1 INTRODUÇÃO

O presente trabalho tem como objetivo apresentar uma análise estatística por meio de modelos de regressão linear para estimar de maneira parcimoniosa a estatura de indivíduos adultos baseada em gênero, comprimento em milímetro dos pés e comprimento em milímetro das mãos. A análise é feita a partir dos dados da base “*Stature Estimation based on Hand Length and Foot Length*” presente no repositório *Miscellaneous DataSets*, contendo 155 observações de 80 homens e 75 mulheres.

2 MATERIAL E MÉTODOS

2.1 Material

A base de dados selecionada é proveniente do repositório de dados *Miscellaneous DataSets*, contendo 5 variáveis, sendo elas:

Id: Identificação dos indivíduos na base, não sendo considerada no estudo.

Stature: Altura do indivíduo (em milímetros).

Hand length: Comprimento da mão (em milímetros).

Foot length: Comprimento do pé (em milímetros).

Gender: Gênero (1: masculino; 2: feminino)

Pela descrição das variáveis, constata-se a presença de uma variável com duas categorias. Tais variáveis foram introduzidas ao software de análise estatística *R* e analisadas utilizando os pacotes “*MASS*” e “*car*”.

2.2 Métodos

Neste estudo, trabalhou-se com regressão linear múltipla em um primeiro momento devido a quantidade de variáveis presentes e a eventual relação entre a variável resposta e as variáveis auxiliares. Com o seu uso, torna-se possível aferir a relação entre a variável independente com as variáveis regressoras, tendo como finalidade explicar a variação da variável resposta. Contudo, fez-se necessário a inclusão de termos quadráticos no modelo, após a realização da análise de resíduos, obtendo-se:

3 RESULTADOS E DISCUSSÃO

3.1 Análise Descritiva e Exploratória

Inicialmente, foram construídos gráficos de dispersão para observar a relação entre a variável resposta e cada uma das variáveis explicativas, conforme apresentado nas figuras 1, 2 e 3:

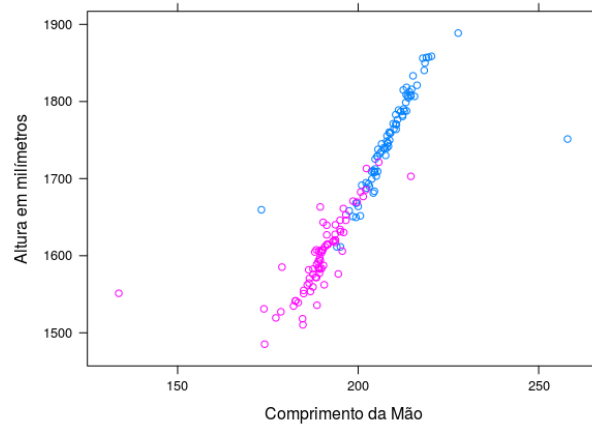


Figura 1: Gráfico de dispersão entre as variáveis *comprimento da mão* e *altura (mm)*.

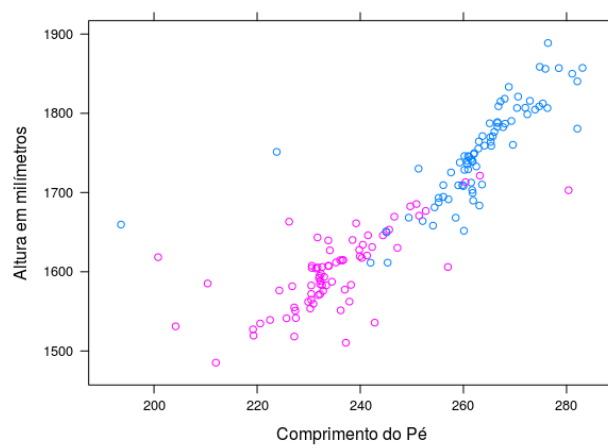


Figura 2: Gráfico de dispersão entre as variáveis *comprimento do pé* e *altura (mm)*.

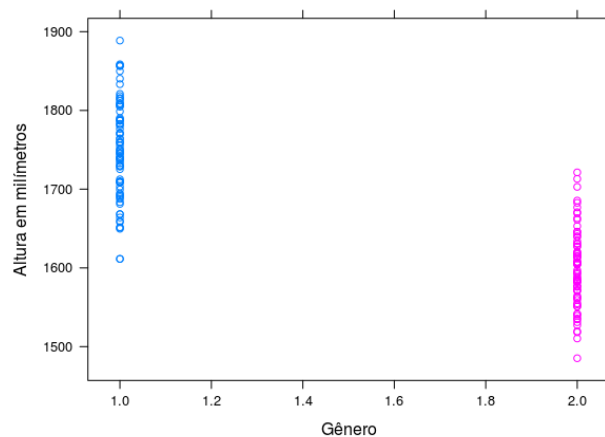


Figura 3: Gráfico de dispersão entre as variáveis *gênero* e *altura (mm)*.

Pelas figuras, constata-se uma relação aparentemente linear e crescente entre as variáveis no caso de comprimento de pé e comprimento de mão, bem como uma visível relação entre o gênero e a altura. A cor azul refere-se aos dados do sexo masculino e a cor rosa do sexo feminino.

Entretanto, por meio da utilização de gráficos de densidade, o aspecto da não normalidade torna-se evidente, sendo decorrente da presença unificada das observações condizentes a variável gênero. As figuras 4, 5 e 6 apresentam os gráficos de densidade de acordo com a altura, comprimento da mão e comprimento do pé. As linhas na cor vermelha e azul, indicam as médias das variáveis diante os indivíduos do sexo masculino e feminino.

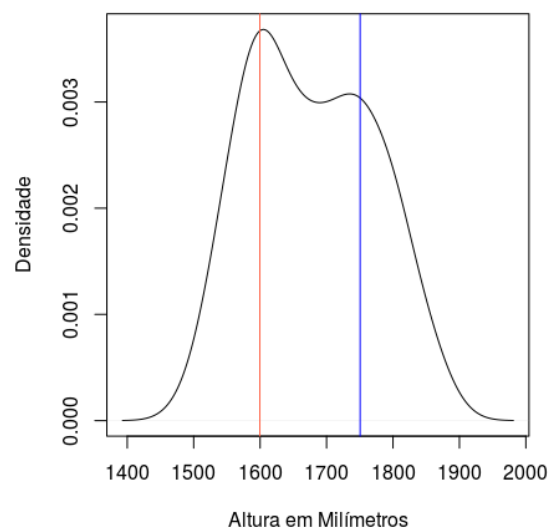


Figura 4: Gráfico de densidade da variável *altura (mm)*

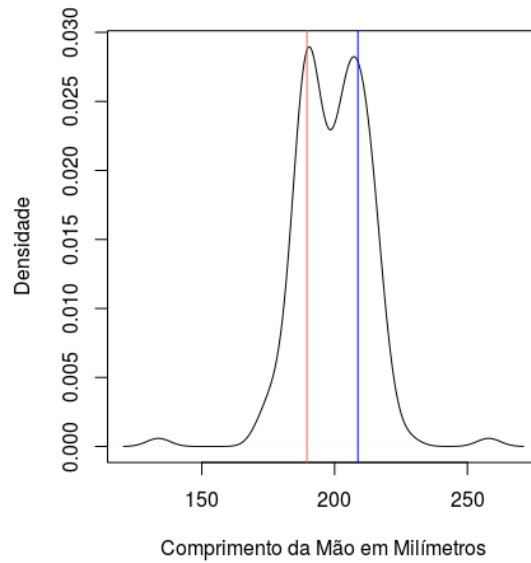


Figura 5: Gráfico de densidade da variável *comprimento da mão (mm)*

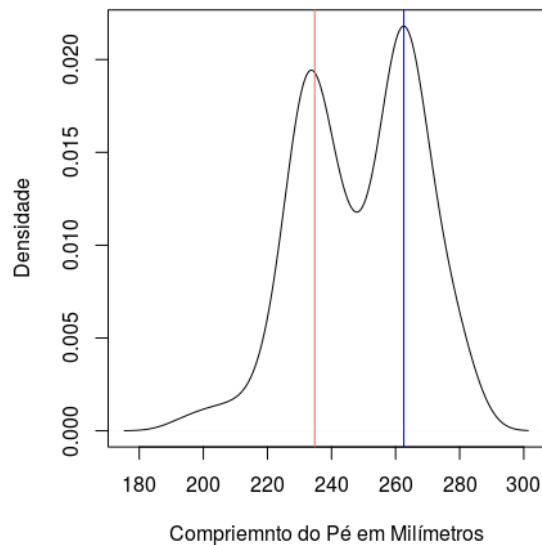


Figura 6: Gráfico de densidade da variável *comprimento do pé (mm)*

3.2 Ajuste do Modelo de Regressão

Considerando a análise descritiva e exploratória, decidiu-se por dividir a modelagem em duas partes. Portanto, foram ajustados modelos diferentes para os dados de indivíduos do sexo feminino e do sexo masculino.

3.2.1 Modelo de Regressão para o gênero feminino

Considerando a variável altura como resposta e as demais variáveis auxiliares, ajustou-se o modelo cuja especificação é descrito abaixo:

$$E[\widehat{y}|x] = \widehat{\beta}_0 + \widehat{\beta}_1 mao + \widehat{\beta}_2 pe$$

Pelo gráfico dos *resíduos vs valores ajustados*, observou-se um comportamento polinomial de segundo grau justificando então a entrada do quadrado das variáveis já presentes no modelo, conforme a figura 7. Além disso, observou-se pelo Teste de Normalidade de Shapiro-Wilk a presença da não normalidade dos dados e, pela *Distância de Cook*, a evidência de candidatos a pontos discrepantes. O coeficiente de determinação múltiplo ajustado foi de 0,6056.

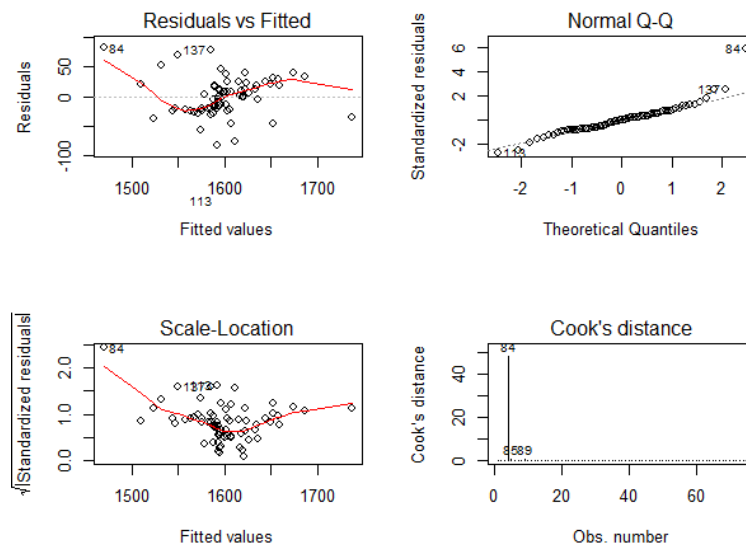


Figura 7: Gráfico de resíduos

Considerando as observações anteriores, foram incluídos termos quadráticos ao modelo, para cada uma das variáveis. As observações 3, 4, 5 e 9 foram retiradas após realizado o diagnóstico de influência com a finalidade de averiguar o impacto dessas observações sobre o modelo. Em seguida foi utilizado o método *backward* para fazer a inclusão das variáveis. De acordo com o método, permaneceram significativos os termos *comprimento da mão*, *comprimento do pé* e *comprimento do pé ao quadrado*, obtendo-se o modelo:

$$E[\widehat{y}|x] = 2376 - 24.83pe + 0.04381pe^2 + 13.86mao$$

A análise de resíduos com base no novo modelo, apresentou homocedasticidade e adequação à distribuição normal (p-valor 0,9239 para o teste de Shapiro-Wilk). O coeficiente de determinação ajustado foi de 0,9189.

De acordo com as estimativas obtidas, pode-se concluir que o intercepto não pode ser interpretado, pois não possui sentido prático. Mantendo-se fixo o valor do comprimento do pé, o incremento na estimativa da altura média de indivíduos do sexo feminino é de 13,86 milímetros para cada milímetro de aumento no comprimento da mão. O incremento na estimativa da altura média de indivíduos do sexo feminino, mantendo-se fixo o comprimento da mão, é de 0,04381 milímetros para cada unidade de milímetro ao quadrado e -24,83 milímetros para cada unidade de milímetro de aumento no comprimento do pé.

3.2.2 Modelo de Regressão para o gênero masculino

Partindo da ideia apresentada no primeiro modelo ajustado considerando o gênero feminino, utilizou-se o modelo ajustado descrito abaixo no caso do gênero masculino:

$$E[\widehat{y}|x] = \widehat{\beta}_0 + \widehat{\beta}_1 mao + \widehat{\beta}_2 pe$$

Na análise de resíduos, foi observada a falta de homocedasticidade e a presença de possíveis pontos influentes. Além disso, observou-se pelo Teste de Normalidade de Shapiro-Wilk a presença da não normalidade dos dados. O coeficiente de determinação ajustado foi de 0,6945.

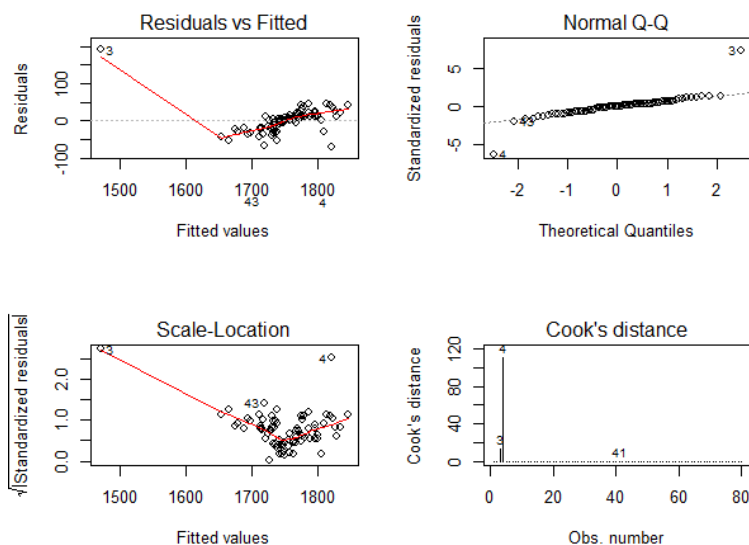


Figura 8: Gráfico de resíduos

Considerando as observações anteriores e conforme a figura 8, foram incluídos termos quadráticos ao modelo, para cada uma das variáveis. As observações 3, 4 e 68 foram retiradas após a realização do diagnóstico de influência. Em seguida foi utilizado o método *backward* para fazer a inclusão das variáveis. De acordo com o método, permaneceram significativos os termos *comprimento da mão ao quadrado*, *comprimento do pé* e *comprimento do pé ao quadrado*, obtendo-se o modelo:

$$E[\widehat{y}|x] = -808 + 0.02385mao^2 + 11.53pe - 0.0218pe^2$$

A nova análise de resíduos mostrou homocedasticidade e adequação à distribuição normal (p-valor 0,1993 para o teste de Shapiro-Wilk). O coeficiente de determinação ajustado foi de 0,9734.

De acordo com as estimativas obtidas, pode-se concluir que o intercepto não pode ser interpretado, pois não possui sentido prático nesta situação. Mantendo-se fixo o valor do comprimento do pé, o incremento na estimativa da altura média do indivíduo do sexo masculino é de 0,02385 milímetros para cada unidade de aumento no comprimento da mão ao quadrado. O incremento na estimativa da altura média, mantendo-se fixo o comprimento da mão, é de 0,0218 milímetros para cada unidade de milímetro ao quadrado e 11,5 milímetros para cada unidade de milímetro de aumento no comprimento do pé.

4 CONSIDERAÇÕES FINAIS

Fixando o gênero (feminino ou masculino), as variáveis *comprimento do pé* e *comprimento da mão* são eficientes para predizer a altura do indivíduo. Todavia, observações femininas com comprimento de mão maiores que 205,7 milímetros ou menores que 173,9 milímetros, assim como comprimentos de pés superiores à 263,2 milímetros ou inferiores à 204,2 milímetros, não podem ser utilizadas no modelo de regressão ajustado para os indivíduos do sexo feminino. Assim como observações masculinas que apresentam comprimentos de pés fora do intervalo de [242; 283,1] milímetros e comprimentos de mão que fogem de [194,1; 220,3] milímetros não podem ser utilizados no modelo de regressão ajustado para indivíduos do gênero do masculino.

Além disso, vale ressaltar que a retirada de algumas observações em ambos os modelos, mostrou uma melhora significativa no ajuste do mesmo. Todavia, investigações subsequentes realizadas junto ao pesquisador podem revelar razões para a retirada das mesmas da análise.