

## Índice de massa corporal de adolescentes e atividade física: um estudo de caso com diferentes abordagens de análise

Helen Maria P. Oliveira<sup>1</sup> †, Eric B. Ferreira<sup>2</sup>

<sup>1</sup>Mestranda em Estatística Aplicada e Biometria, Universidade Federal de Alfenas.

<sup>2</sup> Professor Adjunto IV, ICEx, Universidade Federal de Alfenas.

**Resumo:** *O número de pessoas acima do peso vem crescendo muito nos últimos anos e parte disso de deve à má alimentação da população aliada à falta da prática de exercícios físicos regulares. O trabalho de Costa (2010) apresenta um experimento que relaciona o Índice de Massa Corporal (IMC) com a prática de atividade física de adolescentes em Portugal. O objetivo deste trabalho é apresentar diferentes abordagens analíticas deste mesmo experimento. Tais reflexões são fundamentais no planejamento e consequente análise de experimentos similares. Há indícios para crer que o IMC pode ser diferente entre homens e mulheres e, além disso, como não é foco principal a comparação dos gêneros, talvez seja mais aconselhado analisá-los em dois experimentos distintos. Ainda, a prática de atividade física não interferiu no IMC dos adolescentes.*

**Palavras-chave:** Índice de Massa Corporal; Análise de variância; Modelo aleatório; Componentes de variância.

**Abstract:** *The number of overweight people has been increasing in recent years and part of it due to the poor diet of the population combined with the lack of regular physical exercise. The work of Costa (2010) presents an experiment that relates the Body Mass Index (BMI) with the physical activity of adolescents in Portugal. The objective of this paper is to present different analytical approaches to the same experiment. Such reflections are fundamental in planning and consequent analysis of similar experiments. There is evidence to believe that BMI may be different between men and women and, in this, as it is not the main focus comparison of genres, it may be more desirable to analyze them in two separate experiments. Still, the practice of physical activity did not affect the BMI of adolescents.*

**Keywords:** Body Mass Index; Analysis of variance; random model; Variance components.

## Introdução

Atualmente, o número de pessoas que estão acima do peso vem crescendo consideravelmente. Em geral, não é difícil reconhecer a obesidade ou até mesmo o sobrepeso, mas o diagnóstico correto requer que se identifiquem os níveis de risco, o que, frequentemente, necessita de algumas formas de quantificação. O Ministério da Saúde considera “acima do peso” as pessoas com um Índice de Massa Corporal (IMC) maior que 25. O IMC é calculado dividindo-se o peso pela altura ao quadrado. Acima de 30, a pessoa ser considerada “obesa” (PASSARINHO, 2012).

Segundo o Ministério da Saúde, quase metade da população brasileira sofre com este problema. De acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE), o sobrepeso atinge mais de 30% das crianças entre 5 e 9 anos de idade, cerca de 20% da população entre 10 e 19 anos e nada menos que 48% das mulheres e 50,1% dos homens acima de 20 anos. Mantido o ritmo atual de crescimento do número de pessoas acima do peso, em dez anos 30% da população será obesa, resultando em um sério problema de saúde pública, já que o sobrepeso impulsiona o aparecimento de doenças crônicas, como o diabetes e doenças do coração (SOARES; RITTO, 2010).

---

† Autor correspondente: [helen.pedrosa@yahoo.com.br](mailto:helen.pedrosa@yahoo.com.br).

Apesar de a obesidade estar relacionada a fatores genéticos, há importante influência do sedentarismo e de padrões alimentares inadequados no aumento dos índices brasileiros. A prática regular de exercícios físicos está longe de fazer parte dos hábitos dos mesmos. Soares e Ritto (2010) afirmam que uma pesquisa de 2008 do IBGE mostrou que apenas 10,2% da população com 14 anos ou mais tem alguma atividade física regular.

O tratamento da obesidade é complexo e multidisciplinar. Não existe nenhum tratamento farmacológico em longo prazo que não envolva a mudança de estilo de vida. A obesidade é um desafio mundial, pelo que representa de redução na expectativa de vida e nos custos dos serviços de saúde (ABESO, 2009).

Assim, o objetivo deste trabalho é discutir diferentes abordagens analíticas de um experimento desbalanceado cujo foco é a relação existente entre o índice de massa corporal e a atividade física de adolescentes.

## Metodologia

Este trabalho tomou como base os dados que podem ser encontrados no trabalho de Costa (2010). Os dados são oriundos de uma pesquisa feita entre 2008 e 2009 com 299 alunos de 15 à 19 anos de quatro escolas de Ensino Médio de Portugal: E.S. Homem Cristo, E.S. José Estêvão, E.S. Dr. Jaime Magalhães Lima e E.S. Mário Sacramento. Os alunos foram divididos em grupos diferentes, de acordo com o gênero (feminino e masculino), a prática de atividade física em anos anteriores (sim e não) e a prática de atividade física atualmente (sim e não), que resultou na combinação de um fatorial triplo com dois níveis: **FSS**: mulheres que exercitaram a prática de exercícios físicos em anos anteriores à pesquisa e atualmente; **FSN**: mulheres que exercitaram a prática de exercícios físicos em anos anteriores à pesquisa, mas não atualmente; **FNS**: mulheres que exercitaram a prática de exercícios físicos atualmente, mas não em anos anteriores; **FNN**: mulheres que nunca exercitaram a prática de exercícios físicos; **MSS**: homens que exercitaram a prática de exercícios físicos em anos anteriores à pesquisa e atualmente; **MSN**: homens que exercitaram a prática de exercícios físicos em anos anteriores à pesquisa, mas não atualmente; **MNS**: homens que exercitaram a prática de exercícios físicos atualmente, mas não em anos anteriores; **MNN**: homens que nunca exercitaram a prática de exercícios físicos.

Após a divisão dos grupos, foi feita a medição do Índice de Massa Corporal (IMC), com o objetivo de avaliar o peso corporal dos jovens. Este índice é definido pelo quociente entre o peso (em quilogramas) e o quadrado da altura (em metros), ou seja:

$$IMC = \frac{Peso}{Altura^2}$$

O valor assim obtido estabelece o diagnóstico da obesidade e caracteriza também os riscos associados. A Tabela 1 traz a interpretação do IMC segundo a Organização Mundial da Saúde (OMS, 2000).

Tabela 1: Tabela dos graus de risco e do tipo de obesidade em função do IMC.

IMC ( $Kg/m^2$ )	Grau de Risco	Tipo de Obesidade
$\leq 18$	Magreza	Ausente
]18; 25]	Peso Saudável	Ausente
]25; 30]	Moderado	Sobrecarga ponderal (Pré-Obesidade)
]30; 35]	Alto	Obesidade de grau I
]35; 40[	Muito Alto	Obesidade de grau II
$\geq 40$	Extremo	Obesidade de grau III (Mórbida)

Fonte: Adaptado de OMS (2000).

Os dados de Costa (2010), que serão utilizados neste trabalho, segundo a classificação trazida pela Tabela 1, apresentam 10,5% de adolescentes abaixo do peso normal, quase 80% de pessoas no peso

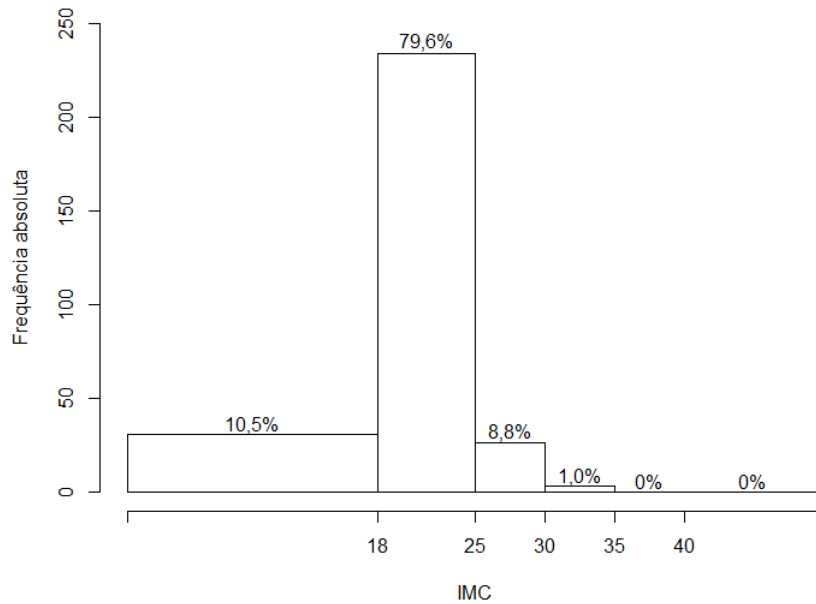


Figura 1: Distribuição de frequência dos valores amostrais de IMC segundo sua classificação em categorias.

saudável, quase 10% de pessoas com sobrepeso e 1% de adolescentes com obesidade de grau I, conforme pode se ver na Figura 1.

Todas as análises e figuras foram feitas no software R (R CORE TEAM, 2014). Em particular, as análises de variância foram feitas por meio do pacote `ExpDes.pt` (FERREIRA; CAVALCANTI; NOGUEIRA, 2013).

## Resultados e Discussão

Como primeira atitude perante os dados, foi realizada uma análise exploratória. A Tabela 2 explicita o desbalanceamento previamente conhecido. Foi computado o índice de desbalanceamento, utilizado por Martins Jr., Ferreira e Ramos (2013),  $\delta = \frac{\max(J_i)}{\min(J_i)}$ , em que  $J_i$ , com  $i = 1, \dots, 8$  é o número de repetições do  $i$ -ésimo tratamento. Para a presente amostra,  $\delta = 6,44$ ; o que sugere um grande desbalanceamento, uma vez que esses mesmos autores, em seu estudo de simulação, trabalharam com índices variando entre 1,17 e 8.

Tabela 2: Número de repetições por tratamento.

Tratamentos	FNN	FNS	FSN	FSS	MNN	MNS	MSN	MSS
Número de repetições	43	58	17	58	13	20	9	81

Em seguida, foram feitos box-plots para os tratamentos a fim de detectar possíveis valores discrepantes. A Figura 2(a) sugere a presença de 1, 3 e 1 *outliers*, respectivamente nos tratamentos FNS, FSS e MSS. Ressalta-se que os três tratamentos que apresentaram valores discrepantes receberam resposta “sim” à pergunta sobre a prática de atividade física atualmente. Após a remoção dos valores discrepantes, a Figura 2(b) ainda apresenta indícios de que pode haver diferença entre as medianas dos tratamentos e que as variâncias podem não ser homogêneas.

Então, a distribuição dos IMCs ao longo dos tratamentos começou a ser investigada por meio de histogramas (Figura 3). A primeira linha de histogramas da Figura 3 refere-se a IMCs do sexo feminino, e a segunda linha ao sexo masculino. Chamam a atenção a distribuição quase uniforme do tratamento MNN, sugerindo que para os meninos que não praticam e não praticaram atividade física, existem os

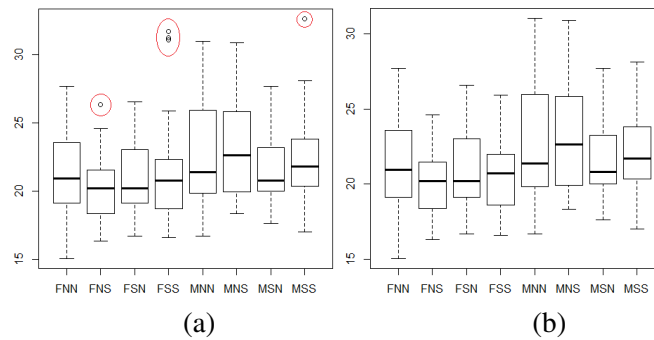


Figura 2: *Box-plots* dos dados de IMC para cada tratamento, antes (a) e após (b) a remoção das observações discrepantes.

mais diversos IMCs; e que os maiores IMCs (> 30) que podem ser vistos primordialmente no sexo masculino, nos tratamentos MNN, MNS, MSS e FSS.

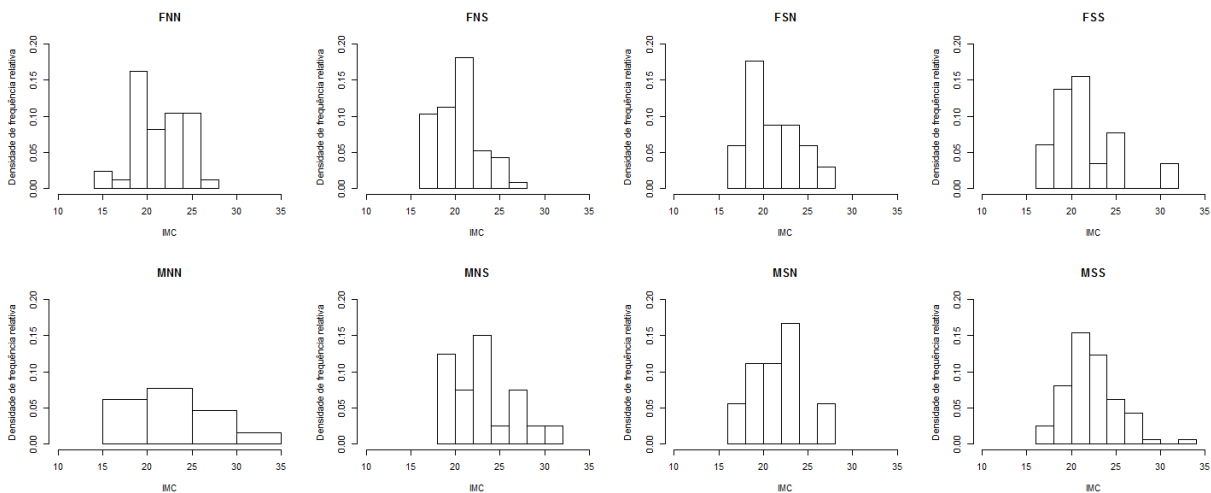


Figura 3: Histogramas dos IMCs para cada um dos tratamentos, demonstrando as distribuições de frequências.

Executaram-se testes de normalidade para todos os tratamentos por meio do teste de Shapiro-Wilk à 5% de significância, mostrando que para nenhum dos tratamentos deve-se rejeitar a hipótese nula de normalidade (Tabela 3).

Tabela 3: Teste de normalidade Shapiro-Wilk, a 5% de significância, por tratamento.

Tratamentos	FNN	FNS	FSN	FSS	MNN	MNS	MSN	MSS
Repetições	43	58	17	58	13	20	9	81
Valor-p	0,3662	0,0928	0,5161	0,0578	0,6102	0,1466	0,6306	0,1698

Em seguida procedeu-se a análise de variância considerando um modelo fixo em delineamento inteiramente casualizado (DIC). Com seus resíduos, foram verificadas as pressuposições de normalidade, homogeneidade de variâncias e independência, utilizando-se os testes de Shapiro-Wilk ( $p=0,015$ ), Bartlett ( $p=0,4180$ ) e Durbin-Watson ( $p=0,1456$ ), respectivamente, todos a 5% de significância. Nota-se que, embora os tratamentos sejam individualmente normais, os resíduos do modelo fixo não podem ser considerados como tal. Por outro lado, não se pode dizer as pressuposições de homoscedasticidade e independência foram violadas. A Tabela 4 traz os valores-p dos referidos testes.

Tabela 4: Valores-p dos testes de normalidade, homogeneidade e independência dos resíduos do modelo fixo em DIC.

Testes	Shapiro-Wilk	Bartlett	Durbin-Watson
Valor-p	0,0015	0,4180	0,1456

O teste F da ANAVA acusa a diferença entre médias ( $p=0,0013$ ), tendo o experimento um coeficiente de variação baixo (13,18%). Simultaneamente foi realizado o teste não paramétrico de Kruskal-Wallis. Este também apontou que há diferença entre os tratamentos ( $p=0,0056$ ).

Então partiu-se para os testes de comparações múltiplas de médias. Foram realizados os testes de Tukey, Student-Newman-Keuls (SNK), Bonferroni (Bonf), Bootstrap, Scott-Knott (SK), t de Student (LSD), Duncan e Calinski & Corsten bootstrap (CCF), sempre considerando-se 5% de significância. Todos concordaram com o teste F, exceto o teste de Bootstrap de Ramos e Ferreira (2009), que se mostrou conservador (Tabela 5).

Tabela 5: Testes de comparações múltiplas de média de Tukey, Student-Newman-Keuls (SNK), Bonferroni (Bonf), Bootstrap, Scott-Knott (SK), t de Student (LSD), Duncan e Calinski &amp; Corsten bootstrap (CCF)

Tratamentos	Médias	TuKey	SNK	Bonf	Bootstrap	SK	LSD	Duncan	CCF
MNS	22,90	a	a	a	a	a	a	a	a
MNN	22,53	ab	ab	ab	a	a	ab	ab	a
MSS	22,12	ab	ab	ab	a	a	ab	abc	a
MSN	21,73	ab	ab	ab	a	a	abc	abc	a
FSS	21,41	ab	ab	ab	a	b	bc	abc	b
FNN	21,22	ab	ab	ab	a	b	bc	bc	b
FSN	20,97	ab	ab	ab	a	b	bc	bc	b
FNS	20,29	b	b	b	a	b	c	c	b

Outra informação importante trazida pela Tabela 5 é que a ordenação das médias mostra uma clara divisão entre mulheres e homens, com os maiores IMCs sendo os masculinos.

Com isso, uma outra abordagem - talvez ainda mais útil - seja a partição da fonte de variação tratamentos em três novas fontes: “gênero”, “atividade física atual” e “atividade física anterior”. Ainda considerando um modelo fixo em DIC, mas dessa vez em esquema fatorial triplo, uma nova análise da variância foi feita (Tabela 6).

Tabela 6: Análise da variância do experimento em esquema fatorial triplo em DIC (modelo fixo).

Fonte de Variação	GL	SQ	QM	Fc	p-valor
Gênero	1	123,58	123,58	13,17	0,0003
Anteriores	1	1,67	1,67	0,17	0,6727
Atual	1	3,09	3,09	0,32	0,5662
Gênero*Anteriores	1	25,32	25,32	2,70	0,1014
Gênero*Atual	1	7,73	7,73	0,82	0,3644
Anteriores*Atual	1	11,03	11,03	1,17	0,2790
Gênero*Anteriores*Atual	1	5,02	5,02	0,53	0,4646
Resíduo	291	2728,71	9,37		
Total	298	2906,2	9,75		

Com esse tipo de abordagem conseguimos observar os efeitos simples. De acordo com o teste F, à 5% de significância temos que os fatores prática de exercícios físicos em anos anteriores e atualmente também foram não significativos. Por sua vez, o fator “gênero” se mostrou significativo, apontando que

no gênero feminino a Índice de Massa Corporal é mais baixo (Tabela 7), suspeita já existente desde a abordagem em DIC simples.

Tabela 7: Teste de Tukey para o fator gênero.

Gênero	Média
Masculino	22,26a
Feminino	20,95b

Nesse ponto da análise, resgata-se uma comparação visual entre a distribuição de frequências dos IMCs masculinos e femininos (Figura 4). Note que os IMCs masculinos (em azul) sobrepostos aos femininos (em vermelho) realmente parecem ser deslocados ligeiramente para a direita. Além disso, os IMCs femininos apresentam maior dispersão (amplitude); 16,05 contra 14,31 de amplitude dos homens.

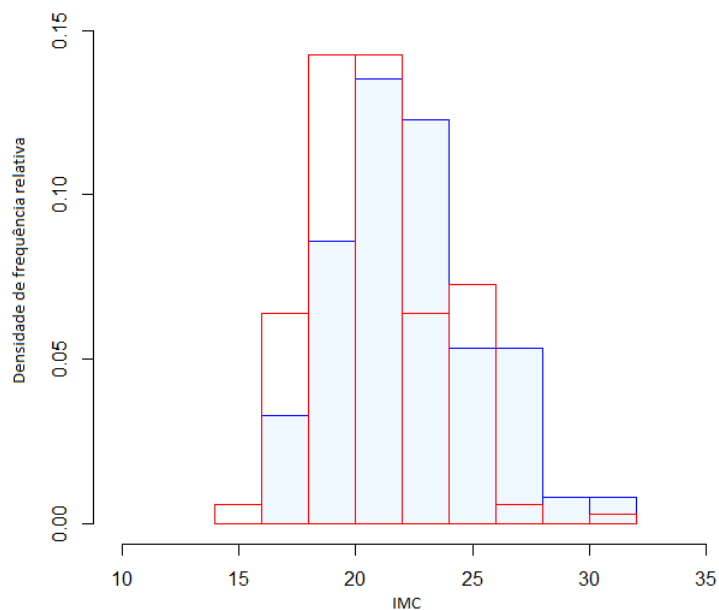


Figura 4: Histogramas dos IMCs masculinos (azul) e femininos (vermelho).

Como abordagem analítica final - e mais correta - vamos considerar que os dados do experimento foram obtidos através de entrevistas. Assim os tratamentos foram formados por componentes aleatórios. Desta forma, o esquema fatorial com modelo de efeitos fixos não é o mais indicado. Ao responder seu gênero e sua prática de atividades físicas atualmente e anteriormente, os respondentes determinavam a qual tratamento pertenceriam. Essa prática gerava intrinsecamente um experimento desbalanceado e de efeitos aleatórios. Neste caso, deve-se utilizar o esquema fatorial triplo aleatório, em DIC.

A Tabela 8 mostra os componentes de variância do modelo aleatório em esquema fatorial triplo, em DIC, obtidos via algoritmo de Ricks:

$$y_{ijk} = \mu + a_i + b_j + c_n + ab_{ij} + ac_{in} + bc_{jn} + abc_{ijn} + e_{(in)k}$$

em que  $y_{ijk}$  é a observação da  $k$ -ésima repetição do  $i$ -ésimo gênero, na  $j$ -ésima atividade física atual e  $n$ -ésima atividade física anterior;  $\mu$  é uma constante comum a todas as observações;  $a_i$  é o efeito do  $i$ -ésimo gênero;  $b_j$  é o efeito da  $j$ -ésima atividade física atual;  $c_n$  é o efeito da  $n$ -ésima atividade física anterior;  $ab_{ij}$ ,  $ac_{in}$ ,  $bc_{jn}$ ,  $abc_{ijn}$  são as interações duplas e tripla de efeitos;  $e_{(in)k}$  é o erro aleatório associado a  $y_{ijk}$ .

Dessa forma, ficam evidentes quais os quadrados médios que devem ser utilizados ao se fazer o teste F. Na tabela 8,  $I$  é o número de níveis do fator gênero,  $J$  é o número de níveis do fator prática

de exercícios físicos em anos anteriores,  $N$  é o número de níveis do fator prática de exercícios físicos atualmente e  $K$  é a média aritmética das repetições de todos os tratamentos. Além disso, consideram-se  $a_i \sim N(0, \sigma^2 a)$ ,  $b_j \sim N(0, \sigma^2 b)$ ,  $c_n \sim N(0, \sigma^2 c)$ ,  $ab_{ij} \sim N(0, \sigma^2 ab)$ ,  $ac_{in} \sim N(0, \sigma^2 ac)$ ,  $bc_{jn} \sim N(0, \sigma^2 bc)$ ,  $abc_{ijn} \sim N(0, \sigma^2 abc)$  e  $e_{(ijn)k} \sim N(0, \sigma^2)$ .

Tabela 8: Componentes de variância do modelo aleatório em esquema fatorial triplo, em DIC, obtidos via algoritmo de Ricks

	I	J	N	K	Esperanças dos Quadrados Médios
	a	a	a	a	
	i	j	n	k	
$a_i$	1	J	N	K	$\sigma^2 + K\sigma_{abc}^2 + JK\sigma_{ac}^2 + NK\sigma_{ab}^2 + JNK\sigma_a^2$
$b_j$	I	1	N	K	$\sigma^2 + K\sigma_{abc}^2 + IK\sigma_{bc}^2 + NK\sigma_{ab}^2 + JNK\sigma_b^2$
$c_n$	I	J	1	K	$\sigma^2 + K\sigma_{abc}^2 + IK\sigma_{bc}^2 + JK\sigma_{ac}^2 + IJK\sigma_c^2$
$ab_{ij}$	1	1	N	K	$\sigma^2 + K\sigma_{abc}^2 + NK\sigma_{ab}^2$
$ac_{in}$	1	J	1	K	$\sigma^2 + K\sigma_{abc}^2 + JK\sigma_{ac}^2$
$bc_{jn}$	I	1	1	K	$\sigma^2 + K\sigma_{abc}^2 + IK\sigma_{bc}^2$
$abc_{ijn}$	1	1	1	K	$\sigma^2 + K\sigma_{abc}^2$
$e_{(ijn)k}$	1	1	1	1	$\sigma^2$

Coincidentemente, as decisões de aceitar ou rejeitar as hipóteses nulas foram iguais às do modelo fixo. Entretanto, a partir desse modelo podem-se estimar os componentes da variância:  $\sigma_a^2 = 0,656$ ,  $\sigma_b^2 = -0,176$ ,  $\sigma_c^2 = -0,049$ ,  $\sigma_{ac}^2 = 0,036$ ,  $\sigma_{bc}^2 = 0,08$ ,  $\sigma_{ab}^2 = 0,271$  e  $\sigma_{abc}^2 = -0,11$ , que segundo Gomes (2009) podem acontecer de serem negativas. A análise das estimativas encontradas não será abordada neste artigo, podendo ser alvo de estudos em trabalhos futuros. Mas percebemos a supremacia da variância associada ao gênero, ou seja, este fator tem maior influência sobre a variação dos tratamentos.

## Conclusões

Neste trabalho foi possível discutir diversas abordagens de análise de experimentos desbalanceados que tratam da comparação de índices de massa corporal (IMC). Tais reflexões são fundamentais no planejamento e consequente análise de experimentos similares.

Além disso, pode-se observar uma possível discrepância entre IMCs de adolescentes dos gêneros masculino e feminino. Mesmo sendo uma métrica que pondera o peso pelo quadrado da altura, o IMC talvez não seja capaz de deixar homens e mulheres comparáveis. Por sua vez, a prática atual ou progressiva de atividade física não pareceu interferir no IMC de adolescentes dessa região do Portugal.

## Referências Bibliográficas

ABESO - Associação Brasileira para o Estudo da Obesidade e da Síndrome Metabólica. *Diretrizes brasileiras de obesidade 2009/2010*, 3ª ed. - Itapevi, SP: AC Farmacêutica, 2009.

FERREIRA, E. B., CAVALCANTI, P. P., NOGUEIRA, D. A. *ExpDes.pt: Experimental Designs package* (Portuguese). R package version 1.1.2. 2013.

COSTA, C. P. S. *Um estudo sobre a robustez do teste-F*. Dissertação (Mestrado em Matemática e Aplicações) - Universidade de Aveiro, Portugal. 2010. 114 p.

GOMES, F. P. *Curso de estatística experimental* 15. ed. Piracicaba: FEALQ, 2009.

MARTINS JR., J. M.; FERREIRA, E. B., RAMOS, P. S. Alternatives for correcting the Tukey's statistics for unbalanced experiments. *Sigmae*, Alfenas, v.2, n.2, p. 21-42. 2013.

PASSARINHO, N. *Quase metade da população está acima do peso, diz Saúde*. 2010. Acesso em: 23 jul 2014.

OMS. Organização Mundial da Saúde. *Obesity: preventing and managing the global epidemic*. Report of a World Health Organization Consultation. Geneva: World Health Organization, 2000. p. 256. WHO Obesity Technical Report Series, n. 284.

R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2014.

RAMOS, P. S., FERREIRA, D. F. Agrupamento de medias via bootstrap de populações normais e não-normais, *Revista Ceres*, v.56, p.140-149, 2009.

SOARES, L.; RITTO, C. *Pesquisa do IBGE confirma que obesidade é epidemia no Brasil*. 2010. Acesso em: 22 jul 2014.