Universidade de Brasília - Unb
Instituto de Ciências exatas - IE
Departamento de Estatística - EST

Aplicação de modelos ANCOVA paramétrico e não paramétrico

LUIZ PAULO MORAIS RORIZ

Orientador: Professor Eduardo Freitas da Silva

Brasília
2018

Luiz Paulo Morais Roriz

Aplicação de modelos ANCOVA paramétrico e não paramétrico

Trabalho de conclusão de Curso apresentado à Universidade de Brasília, como exigência parcial à obtenção do título de bacharel em Estatística.

Orientador:

Professor Eduardo Freitas da Silva

Brasília

2018

Em tudo dai graças, porque esta é a vontade de Deus em Cristo Jesus para convosco.

1 Tessalonicenses 5:18

Resumo

Os modelos de análise de covariância paramétrico e não paramétrico são amplamente utilizados em dados pré-teste e pós-teste, coletados antes e após uma ou mais intervenções, possibilitando a observação do comportamento dos dados de acordo com o tempo. A ANCOVA trata o resultado pré-teste como covariável, variável que representa uma fonte de variação incontrolada no estudo. Foram usadas essas análises nesse estudo, para avaliar o impacto do tratamento de intervenção nutricional, com o fim de promover conhecimento e práticas alimentares saudáveis entre adolescentes. Respostas ordinais nos levam à uma análise não paramétrica. Os resultados foram avaliados por meio de testes de hipóteses sob a abordagem da análise de covariância.

Palavras-chave: Pré-teste, pós-teste, ancova, ancova não paramétrica, comparar médias, covariável.

Sumário

1 INTRODUÇÃO E JUSTIFICATIVA 7

2 OBJETIVOS 8

2.1 Objetivo geral 8

2.2 Objetivo específico 8

3 QUESTIONÁRIO 9

4 BANCO DE DADOS 10

4.1 Introdução 10

4.2 Seleção da amostra 10

5 METODOLOGIA 11

5.1 Análise de covariância (ANCOVA) 11

5.1.1 Modelo estatístico 11

5.1.2 Suposições 12

5.1.3 Teste de hipóteses 12

5.1.4 Teste de linearidade 14

5.1.5 Ajuste e comparação das médias 15

5.1.6 Diagnóstico 16

5.2 ANCOVA não paramétrica 17

5.2.1 Teste de Quade 17

5.2.1.1 Delineamento 17

5.2.1.2 Suposições 18

5.2.1.3 Teste de hipóteses 19

5.2.1.4 Múltiplas comparações 20

5.2.2 Análise Paramétrica de Covariância nos ranks. 20

6 RESULTADOS. 21

6.1 Análise descritiva 21

6.2 Análise de Covariância 27

6.2.1 Paramétrica 27

6.2.2 Análise de Covariância Não paramétrica 37

7 CONSIDERAÇÕES FINAIS. 47

REFERÊNCIAS BIBLIOGRÁFICAS 48

ANEXO

1 Introdução e Justificativa

Em estudos diversos, é de interesse do pesquisador comparar médias entre dois ou mais grupos. A relação entre as variáveis do estudo podem ser influenciadas por uma "terceira". Esta última é denominada como covariável ou ainda, variável auxiliar ou variável concomitante, uma variável que pode ser contínua e representa uma fonte de variação que não foi controlada no experimento, e deve ser incluída na análise estatística.

A Análise de Covariância (ANCOVA), é uma extensão da Análise de Variância (ANOVA). Diferentemente da ANOVA, a ANCOVA pode incluir uma ou mais covariáveis. Desse modo, para utilização desse modelo é preciso verificar se a covariável está linearmente relacionada à variável resposta.

Essa análise combina regressão com análise de variância, e pode ser pensada como uma técnica de controle estatístico, porque a ANCOVA, em análises pré-teste pós-avaliação, trata o resultado pré-teste como covariável, e acredita-se que afete o valor esperado da pontuação do pós-teste (Kirk, 1982).

Esses modelos são amplamente utilizados em estudos clínicos. Os dados são coletados antes e após uma ou mais intervenções, possibilitando a observação do comportamento dos dados de acordo com o tempo.

Controlando a covariável, diminuiremos parte da variância residual, otimizando o modelo estatístico, além de controlarmos uma fonte de variação que afeta a variável dependente. É possível fazer também o ajuste das médias dos tratamentos em função da covariável e, em alguns casos, estimar observações perdidas durante o experimento.

Com as devidas suposições aferidas, normalidade para a variável resposta, homogeneidade de variância, igualdade dos coeficientes de regressão da população dentro dos grupos, e a, já mencionada, relação de linearidade entre covariável e variável resposta, podemos usufruir dessa técnica estatística para obtenção dos nossos resultados.

Já a análise de covariância não paramétrica é necessária uma vez que as suposições do modelo paramétrico são violadas. O procedimento de Quade (1976) consiste em, primeiramente, transformar os escores do pré-teste e do pós-teste, classificando-os por ranking. As classificações das pontuações são então convertidas em pontuações de desvio de grau médio, subtraindo a classificação média de cada observação, formando um escore de classificação do desvio residual. Uma análise de variância é então realizada nesse escores. Basicamente, o procedimento de Quade é calcular uma regressão linear nas classificações corrigidas médias, depois sujeitar os resíduos a uma análise de variância.

Outra forma possível de lidar com essa problemática, poderiam ser os métodos de Puri e Sem (1969), também utilizando análise de covariância não paramétrico.

2 Objetivos

2.1 Objetivo geral

Estudo e emprego de modelos de análise de covariância paramétrico e não paramétrico e suas aplicações.

2.2 Objetivo específico

Serão empregados modelos de Análise de Covariância, em delineamentos, do tipo pré e pós tratamento, na qual a covariável será a medida obtida no pré tratamento.

Pretende-se aplicar essa metodologia em um estudo experimental realizado com adolescentes alunos de sete escolas da rede pública de ensino de Brasília, Distrito Federal, em relação aos seus conhecimentos e hábitos alimentares frequentes. Os adolescentes selecionados, envoltos numa faixa etária de treze a dezoito anos, foram inicialmente randomizados em dois grupos (controle e experimental) e submetidos a um questionário composto por perguntas de ordem quantitativa, e qualitativa ordinal.

Transcorrida essa primeira etapa, os alunos do grupo experimental foram sujeitos a três intervenções. Tratamento este que consistiu em capacitá-los a entender os princípios da alimentação saudável, classificação dos alimentos, importância da leitura dos rótulos e do olhar crítico para propagandas de alimentos, através de palestras e estratégias pedagógicas. Já o grupo controle não sofreu nenhuma intervenção. Após um período de tempo, ambos serão novamente questionados e reavaliados sob o mesmo questionário.

Entende-se que, somente a aplicação do questionário, gere alguma influência nos hábitos dos alunos. Portanto, espera-se que as primeiras respostas atuem como covariável influente nas respostas futuras. Implementaremos o modelo não paramétrico às respostas de natureza ordinal, transformando os escores em rankings.

3 Questionário

O questionário é fracionado em 5 partes: identificação e variáveis sociodemográficas, avaliação do consumo alimentar, avaliação das práticas alimentares, autopercepção e conhecimento. Incialmente são apenas dados de registro e de controle do aluno, como nome, sexo, idade, peso, altura, colégio, turma, participação nas intervenções, além de dados sociodemográficos.

Depois, para avaliar o consumo habitual, o questionário aborda algumas perguntas sobre a alimentação de cada aluno dividido em duas partes, a primeira em relação ao consumo alimentar no dia anterior à realização do questionário e a outra sobre o consumo da última semana. Aborda principalmente alimentos marcadores de uma alimentação saudável e não saudável.

Mais tarde algumas perguntas sobre o comportamento do aluno durante as refeições, ou alguns hábitos alimentares no geral, incluindo perguntas sobre a realização de almoço ou jantar junto ao pai, à mãe, ou responsável, hábito de comer enquanto assiste televisão ou estuda e o hábito de consumir café da manhã, além de questionamentos sobre a autopercepção que cada indivíduo tem sobre a sua própria alimentação.

Por fim, alguns posicionamentos para verificarmos o nível de conhecimento do aluno. Formularam-se questões que englobam conceitos básicos e terminologias sobre alimentação saudável. É apresentado em anexo o modelo final do questionário.

4 Banco de dados

4.1 Introdução

A base de dados usada para ilustrar os modelos propostos, foram coletados para defesa da dissertação da discente Laís Gomes Fonseca de mestrado em nutrição humana da Universidade de Brasília, orientada pela professora Maria Natacha Toral Bertolin, no segundo semestre de 2017. A pesquisa abordou adolescentes entre 13 e 18 anos estudantes do 9º ano de sete diferentes escolas públicas de Brasília da Regional do Plano Piloto e Cruzeiro, foram elas: CEF polivalente, CEF 214 sul, CEF 102 norte, CEF 06 lago sul, CEF 410 norte e CEF caseb.

4.2 Seleção da amostra

Num total de 590 respondentes, 461 adolescentes foram selecionados e devidamente questionados e acompanhados durante todo o estudo, superando os critérios de exclusão. Foram incluídos alunos de ambos os sexos, de qualquer idade que estavam devidamente matriculados no 9º ano do ensino fundamental. Depois à randomização foi feita por escolas para dividir os alunos entre grupos de controle e tratamento, de forma que todas possuíam a mesma chance no sorteio.

Possuímos um total de 131072 observações distribuídas nas 461 unidades amostrais em 88 variáveis (sendo apenas duas do tipo "caractere" e as outras do tipo numérica). Parte das variáveis, exatamente 12, são de domínio identificativo e outras futuramente serão agrupadas.

Algumas variáveis são de grande interesse e devem ser consideradas. Estar presente ou não em cada uma das intervenções apresenta um resultado significativo no estudo, por isso denotamos P1int para o aluno que esteve presente em todas 3 diferentes intervenções.

O instrumento de coleta consistiu de um questionário fracionado em 5 partes, foram elas: identificação e variáveis sociodemográficas, avaliação do consumo alimentar, avaliação das práticas alimentares, autopercepção e conhecimento, compostos por perguntas de ordem quantitativa e qualitativa ordinal.

Na quinta parte do questionário, foram aplicadas questões sobre conhecimento de uma alimentação saudável. Como opções de respostas foram colocadas três categorias, sendo elas: "verdadeiro", "falso" e "não sei", onde a cada resposta correta atribuía-se 1 ponto para o aluno, e para cada resposta incorreta ou desconhecida, o aluno não somava pontos (o gabarito para as questões em evidência se encontram em anexo). Foram 13 questões no total, agrupamos todas elas na variável Nfin (nota final, após cada aplicação do questionário), portanto, a variável assumia valores de 0 a 13.

As questões 15, 29, 30, 31 e 32 geraram variáveis que consideramos no estudo não paramétrico por não apresentarem respostas nominais.

5 Metodologia

5.1 Análise de Covariância (ANCOVA)

Ao tratarmos com um experimento onde a variável resposta está linearmente relacionada a uma variável X que não pôde ser controlada pelo pesquisador, mas foi observada conjuntamente à resposta, denominamos X como covariável e sujeitamos o estudo a uma análise de covariância. Considerada como uma análise útil para diminuir o viés e melhorar a precisão de um experimento, ela testa as principais interações e efeitos de variáveis categóricas na variável resposta, controlando os efeitos de outras variáveis selecionadas, que covariam com a dependente. Caso esse ajuste não seja feito, a covariável pode inflar o erro quadrático médio.

A ANCOVA combina as características da análise de variância com regressão. O princípio básico é a inclusão da covariável no modelo de análise de variância, com a intenção de reduzir a variância dos erros para melhorar a precisão na análise dos efeitos do tratamento. Ela avalia o efeito de uma ou mais covariáveis na variável resposta dividindo o termo erro em dois componentes, um por variação não explicada ou residual, e outro devido a regressão dos escores do pré-teste sobre os escores do pós-teste, logo comparando ao modelo de análise de variância, em um experimento casualizado, o erro quadrático médio é consideravelmente menor.

São empregados, nos modelos de análise de covariância, delineamentos do tipo pré e pós tratamento, na qual a covariável é a medida obtida no pré tratamento. Com isso garantimos que ela não é afetada pelo tratamento e, portanto, pode ser considerada covariável. Entende-se que ela não foi controlada pelo pesquisador e pode afetar no pós tratamento. A análise estatística testa se o efeito dos tratamentos é significativo no pós-teste ajustado pelo pré-teste.

5.1.1 Modelo estatístico

O modelo linear da análise de covariância é:

$$y_{ij} = \mu + \tau_i + \beta_w\left(X_{ij} - \bar{X}_{..}\right) + \epsilon_{ij} \qquad \begin{cases} i = 1, 2 \dots a \\ j = 1, 2 \dots n \end{cases} \tag{1}$$

Onde:

$y_{ij} = valor\ observado\ da\ variável\ resposta$

$\mu = média\ populacional;$

$\tau_i = efeito\ do\ i - ésimo\ tratamento;$

$X_{ij} = valor\ observado\ da\ covariável;$

$\bar{X}_{..} = média\ da\ covariável;$

$\beta_w = coeficiente\ de\ regressão\ linear\ entre\ a\ covariável\ e\ a\ variável\ resposta;$

$\epsilon = erro\ da\ j - ésima\ observação\ no\ grupo\ i.$

5.1.2 Suposições

Para o devido modelo supomos:

- Correlação de linearidade entre covariável e variável resposta ($\beta \neq 0$);
- Os erros são independentes e normalmente distribuídos com média zero e variância comum constante ($e_{ij}\ i.i.d. \sim N(0, \sigma^2)$)
- Os coeficientes de regressão da população dentro dos grupos são iguais, isto é $\beta_1 = \beta_2 = \cdots = \beta_w$.

5.1.3 Teste de hipóteses

Com as devidas suposições aferidas verificamos, por meio de testes de hipóteses a significância dos tratamentos. Construímos a hipótese nula, segundo MONTGOMERY afirmando que o escore médio pós-teste em cada grupo é igual, ajustados aos valores do pré-teste, já na hipótese alternativa assumimos diferenças nos escores médios de pelo menos dois grupos. As hipóteses a qual deseja-se testar é denota por:

H_0: $\tau_1 = \tau_2$

H_1: $pelo\ menos\ um\ \tau_i \neq 0$

A tabela1 mostra a análise de covariância com uma covariável:

					Ajustado pela regressão		
F.V.	G.L.	SQX	SQXY	SQY	Y	G.L.	Q.M.
Tratamentos	$a-1$	T_{xx}	T_{xy}	T_{yy}			
Erro	$a(n-1)$	E_{xx}	E_{xy}	E_{yy}	SSE	$a(n-1)-1$	MSE
Total	$an-1$	S_{xx}	S_{xy}	S_{yy}	SSE^*	$an-2$	
Trat. ajustados					$SSE^* - SSE$	$a-1$	$\frac{SSE^* - SSE}{a-1}$

Tabela 1 - ANCOVA

Onde:

$$T_{xx} = n\sum_{i=1}^{a}(\bar{X}_{i.} - \bar{X}_{..})^2 = \frac{1}{n}\sum_{i=1}^{a} X_{i.}^2 - \frac{X_{..}^2}{an} \tag{2}$$

$$T_{xy} = n\sum_{i=1}^{a}(\bar{X}_{i.} - \bar{X}_{..})(\bar{Y}_{i.} - \bar{Y}_{..}) = \frac{1}{n}\sum_{i=i}^{a}(X_{i.})(Y_{i.}) - \frac{(X_{..})(Y_{..})}{an} \tag{3}$$

$$T_{yy} = n\sum_{i=1}^{a}(\bar{Y}_{i.} - \bar{Y}_{..})^2 = \frac{1}{n}\sum_{i=1}^{a} Y_{i.}^2 - \frac{Y_{..}^2}{an} \quad (4)$$

$$S_{xx} = \sum_{i=1}^{a}\sum_{j=1}^{n}\left(X_{ij} - \bar{X}_{..}\right)^2 = \sum_{i=1}^{a}\sum_{j=1}^{n} X_{ij}^2 - \frac{X_{..}^2}{an} \quad (5)$$

$$S_{xy} = \sum_{i=1}^{a}\sum_{j=1}^{n}\left(X_{ij} - \bar{X}_{..}\right)\left(Y_{ij} - \bar{Y}_{..}\right) = \sum_{i=1}^{a}\sum_{j=1}^{n} X_{ij}Y_{ij} - \frac{(X_{..})(Y_{..})}{an} \quad (6)$$

$$S_{yy} = \sum_{i=1}^{a}\sum_{j=1}^{n}\left(Y_{ij} - \bar{Y}_{..}\right)^2 = \sum_{i=1}^{a}\sum_{j=1}^{n} Y_{ij}^2 - \frac{Y_{..}^2}{an} \quad (7)$$

$$E_{xx} = \sum_{i=1}^{a}\sum_{j=1}^{n}\left(X_{ij} - \bar{X}_{i.}\right)^2 = S_{xx} - T_{xx} \quad (8)$$

$$E_{xy} = \sum_{i=1}^{a}\sum_{j=1}^{n}\left(X_{ij} - \bar{X}_{i.}\right)\left(Y_{ij} - \bar{Y}_{i.}\right) = S_{xy} - T_{xy} \quad (9)$$

$$E_{yy} = \sum_{i=1}^{a}\sum_{j=1}^{n}\left(Y_{ij} - \bar{Y}_{i.}\right)^2 = S_{yy} - T_{yy} \quad (10)$$

Os quadrados médios são calculados através da divisão da soma de quadrados pelo respectivo grau de liberdade.

Estima-se os quadrados médios sem o efeito da covariável por:

$$SSE = E_{yy} - \frac{\left(E_{xy}\right)^2}{E_{xx}} \quad (11)$$

$$MSE = \frac{SSE}{a(n-1)-1} \tag{12}$$

$$\beta = \frac{E_{xy}}{E_{xx}} \tag{13}$$

Estima-se os quadrados médios com o efeito da covariável por:

$$SSE^* = S_{yy} - \frac{\left(S_{xy}\right)^2}{S_{xx}} \tag{14}$$

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} \tag{15}$$

Sob a hipótese nula, a estatística do teste tem distribuição $F_{a-1,\ a(n-1)-1}$ e será rejeitada caso $F_t > F_{\alpha,\ a-1,\ a(n-1)-1}$, valor tabelado ao respectivo nível de significância e graus de liberdade. A distribuição F de Snedecor também conhecida como distribuição de Fisher é frequentemente utilizada na inferência estatística para análise da variância. Uma variável aleatória contínua X tem distribuição F com v_1 e v_2 graus de liberdade, denotada por $F_{v_1,\ v_2}$.

A estatística do teste é:

$$F_t = \frac{\dfrac{SSE^* - SSE}{a-1}}{\dfrac{SSE}{a(n-1)-1}} \tag{16}$$

Com ela verificaremos se os efeitos do tratamento, ajustado à covariável são significativos no resultado da variável resposta.

5.1.4 Teste de linearidade

Podemos agora testar a suposição básica, já mencionada, da correlação de linearidade entre covariável e variável resposta, também com um teste de hipóteses, utilizando a estatística F. Definimos a hipótese nula afirmando que não há correlação, e a rejeição implicaria à hipótese

alternativa, onde se define uma correlação entre a covariável e a variável resposta, denotaremos então:

H_0: $\beta = 0$

H_1: $\beta \neq 0$

Logo, a hipótese nula pode ser testada por:

$$F_\beta = \frac{\frac{(E_{xy})^2}{E_{xx}}}{MSE} \sim F_{1,\ \alpha(n-1)-1} \tag{17}$$

Será rejeitada caso $F_\beta > F_{\alpha,\ a,\ a(n-1)-1}$,

5.1.5 Ajuste e comparação das médias

Para Montgomery, além de testar a hipótese de que não há diferenças nos efeitos do tratamento, frequentemente acha-se útil interpretar os dados para apresentar as médias de tratamento ajustadas. Estas médias ajustadas em função da covariável são calculadas de acordo com:

$$\bar{y}_{i.\ ajust} = \bar{y}_{i.} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..}) \qquad i = 1,2, \dots, a \tag{18}$$

O ajuste da variável obserada Y, pode ser entendido como:

$$y_{ij} - \hat{\beta}\left(X_{ij} - \bar{X}_{..}\right) = \mu + \tau_i + \epsilon_{ij} \tag{19}$$

O erro padrão de qualquer média de tratamento ajustada é:

$$S_{ajust\ yi.} = \left[MSE\left(\frac{1}{n} + \frac{(\bar{x}_{i.} - \bar{x}_{..})}{E_{xx}}\right)\right]^{1/2} \tag{20}$$

Ajustando as medidas, podemos realizar um teste *a posteriori*, chamado teste de Tukey para múltiplas comparações entre as médias. Sônia Vieira, Doutora em Estatística pela USP, afirma que para proceder ao teste de Tukey, é preciso calcular a diferença mínima que deve

haver entre duas médias para que elas possam ser consideradas diferentes ao nível de significância α. Essa diferença é conhecida como diferença mínima significante, o autor do teste, chamou a diferença de "honestly significant difference" (HSD), ou seja, diferença honestamente significante.

A estatística do teste é então definida como:

$$T_\alpha = q(a, f)\sqrt{\frac{MSE}{n}} \tag{21}$$

Caso a diferença entre as médias ajustadas seja menor que T_α rejeitamos a hipótese nula que afirma a igualdade entre as médias, e o ajuste pela covariável teve efeito significativo. Testamos as hipóteses:

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

5.1.6 Diagnóstico

A verificação diagnóstica do modelo de covariância é baseada na análise de resíduos. Para o modelo de covariância, os resíduos são:

$$\epsilon_{ij} = y_{ij} - \hat{y}_{ij} \tag{22}$$

Os valores ajustados serão:

$$\begin{aligned} \hat{y}_{ij} &= \hat{\mu} + \hat{\tau}_i + \hat{\beta}\left(x_{ij} - \bar{x}_{..}\right) \\ &= \bar{y}_{..} + \left[\bar{y}_{i.} - \bar{y}_{..} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..})\right] + \hat{\beta}\left(x_{ij} - \bar{x}_{..}\right) \\ &= \bar{y}_{i.} + \hat{\beta}\left(x_{ij} - \bar{x}_{..}\right) \end{aligned} \tag{23}$$

Então:

$$\epsilon_{ij} = y_{ij} - \bar{y}_{i.} + \hat{\beta}\left(x_{ij} - \bar{x}_{..}\right) \tag{24}$$

5.2 ANCOVA não paramétrica

Devido à violação nas suposições do modelo paramétrico, a análise de covariância não paramétrica se faz necessária. Perguntas como: “Você costuma tomar o café da manhã?” ou “Ontem, quantas vezes você comeu frutas frescas?” geram repostas ordinais e por isso devem ser manipuladas.

Seaman, Algina e Olejnik (1985) afirmam que “rank ANCOVA” possui taxas de erro Tipo I perto do nível nominal, geralmente é mais poderoso que o ANCOVA paramétrico e pode ser substancialmente mais poderoso, quando as suposições de ANCOVA paramétrica são violadas. Isso está de acordo com o estudo de simulação realizado por Conover e Iman (1982), que demonstraram que o teste de Quade e a ANCOVA paramétrica nos postos geralmente estão em acordo virtual.

Segundo CONOVER, testamos a hipótese nula de meios de tratamento iguais em um desenho de blocos ou grupos completos randomizados, mas usa informações sobre o intervalo dos blocos em relação um ao outro e dá mais peso aos blocos que exibem o maior intervalo. Classificamos os rankings dentro do grupo como no teste de Friedman, um teste similar. Convencionou-se nomear o teste pelo nome do seu inventor, Dana Quade (1972, 1979).

O procedimento de Quade, apesar de antigo, é bastante utilizado e eficaz e consiste em, transformar os escores do pré-teste e do pós-teste, classificando-os por ranking. As classificações das pontuações são então convertidas em pontuações de desvio de grau médio, subtraindo a classificação média de cada observação, formando um escore de classificação do desvio residual.

Nesses escores, é então realizada uma análise de variância. Basicamente, o procedimento consiste em calcular uma regressão linear nas classificações corrigidas médias, depois sujeitar os resíduos a ANOVA. Outra forma possível de lidar com essa problemática, poderiam ser os métodos de Puri e Sem (1969), pode ser uma possível sugestão para trabalhos futuros. Esses testes são projetados para ignorar a suposição de homogeneidade dos coeficientes de regressão.

5.2.1 Teste de Quade

5.2.1.1 Delineamento

Os dados consistem em b blocos, $i = 1, 2, \dots, b$, mutuamente independentes de k tratamentos $X_{i1}, X_{i2}, \dots, X_{ik}$. A variável aleatória X_{ij} está associada ao tratamento j no bloco i. O delineamento é então realizado da seguinte forma.

	Tratamento			
Bloco	1	2	. . .	K
1	X_{11}	X_{12}	. . .	X_{1k}
2	X_{21}	X_{22}	. . .	X_{2k}
3	X_{31}	X_{32}	. . .	X_{3k}
. . .	. . .	. . .	. . .	. . .
b	X_{b1}	X_{b2}	. . .	X_{bk}

Seja $R(X_{ij})$ a classificação, de 1 a k, atribuída a X_{ij} dentro do bloco (linha) i. Classificamos então, para o bloco i, as variáveis aleatórias $X_{i1}, X_{i2}, \dots, X_{ik}$ que são comparadas entre si, sendo a classificação 1 atribuída ao menor valor observado, a classificação 2 à segunda menor e assim por diante. A maior observação no bloco i é então atribuída à k. Para cada bloco são atribuídas as devidas classificações. Usa-se classificações médias em caso de empate.

O próximo passo novamente usa as observações originais X_{ij}. As classificações são atribuídas aos próprios blocos de acordo com o tamanho do intervalo de amostra em cada bloco. Este intervalo é a diferença entre as maiores e as menores observações do mesmo.

$$Faixa\ no\ bloco\ i = \underset{j}{máximo}\{X_{ij}\} - \underset{j}{mínimo}\{X_{ij}\} \tag{25}$$

Para cada bloco existe um intervalo de amostra, atribuimos $rank\ 1$ para o bloco com o menor intervalo, $rank\ 2$ para o segundo menor e assim por diante para o bloco com o maior intervalo, que fica $rank\ b$. Usa-se classificações médias em caso de empate. Seja $Q_1, Q_2, \dots, Q_b$ as classificações atribuídas aos blocos $1, 2, \dots, b$, respectivamente.

Finalmente, o bloco de $rank\ Q_i$ é multiplicado pela diferença entre a classificação dentro do bloco i, $R(X_{ij})$ e a classificação média dentro dos blocos, $(k+1)/2$, para obter o produto S_{ij}, onde

$$S_{ij} = Q_i \left[R(X_{ij}) - \frac{k+1}{2} \right] \tag{26}$$

é uma estatística que representa o tamanho relativo de cada observação dentro do bloco, ajustado para refletir a importância relativa do bloco em que aparece.

Seja S a soma de cada tratamento:

$$S_j = \sum_{i=1}^{b} S_{ij} \qquad j = 1, 2, \dots, k \tag{27}$$

5.2.1.2 Suposições

- Os resultados dentro de um bloco não influenciam os resultados dentro dos outros.
- Dentro de cada bloco, as observações podem ser classificadas de acordo com algum critério de interesse.
- O intervalo da amostra pode ser determinado dentro de cada bloco para que os blocos possam ser classificados.

5.2.1.3 Teste de hipóteses

Para calcularmos a estatística de teste calculamos primeiro o termo

$$A_1 = \sum_{i=1}^{b}\sum_{j=1}^{k} {S_{ij}}^2 \tag{28}$$

onde Sij é dado pela equação 26. Denotado por "soma de quadrados total". Outra forma simplificada seria

$$A_1 = b(b+1)(2b+1)k(k+1)(k-1)/72 \tag{29}$$

Em seguida, calculamos o termo

$$B = \frac{1}{b}\sum_{j=1}^{k} {S_j}^2 \tag{30}$$

onde Sj é dado pela equação 27. Chamado de "soma dos quadrados de tratamento". A estatística do teste é

$$T_1 = \frac{(b-1)B}{A_1 - B} \tag{31}$$

se $A_1 = B$, considere o ponto na região crítica e calcule o p-valor como $\left(\frac{1}{k!}\right)^{b-1}$

Observe que T_1 é a análise bidirecional do teste de variância calculada com base nas pontuações S_{ij}, dadas pela equação 26.

A distribuição exata de T_1 é difícil de encontrar, então a distribuição F é usada como uma aproximação, com $k_1 = k - 1$ e $k_2 = (b-1)(k-1)$.

Hipóteses:

H_0: Cada ranking das variáveis aleatórias dentro de um grupo é igualmente provável (isto é, o tratamento tem efeitos idênticos).

H_1: Pelo menos um dos tratamentos tende a produzir maiores valores observados do que pelo menos um outro tratamento.

Rejeita-se as hipóteses nulas no nível α se T_1 exceder o quantil $1 - \alpha$ da distribuição F com $k_1 = k - 1$ e $k_2 = (b-1)(k-1)$.

5.2.1.4 Múltiplas comparações

Somente se o procedimento anterior resultar na rejeição da hipótese nula, serão feitas comparações múltiplas. Os tratamentos i e j são considerados diferentes se a desigualdade

$$\left|S_i - S_j\right| > t_{1-\alpha/2}\left[\frac{2b(A_1 - B)}{(b-1)(k-1)}\right]^{1/2} \tag{32}$$

for satisfeita, onde S_i, S_j, A_1 e B foram dados anteriormente e onde $t_{1-\alpha/2}$ é obtido da tabela da distribuição t-student, com $(b-1)\ (k-1)$ graus de liberdade. Essa comparação é feita para todos os pares de tratamentos, usando o mesmo α usado no teste do quade.

O método de Quade é um caso especial do método mais geral de Puri e Sen (1969), embora as distribuições aproximadas sejam diferentes. Esses desenvolveram uma estatística de teste, L_n que testa a hipótese que não há efeito do tratamento. L_n é distribuído como uma variável aleatória qui-quadrada.

5.2.2 Análise de covariância paramétrica nos ranks

As classificações dos escores pré-teste e pós-teste convertidas em escores de desvio da classificação média, subtraida da classificação média da observação, são chamadas de d_{pre} e d_{post}, respectivamente. Estimando-se o coeficiente de regressão $\hat{\beta}$, pela regressão linear dos mínimos quadrados, podemos prever d_{post} em função de d_{pre}. O grau de desvio do pós-teste, $\hat{d}_{post}$, é então previsto a partir do grau de desvio do pré-teste por:

$$\hat{d}_{post} = d_{pre}\, x\, \hat{\beta}$$

A pontuação do desvio de pós-teste previsto é então subtraída da classificação de desvio observada do pós-teste para formar uma pontuação de classificação de desvio residual, e_r

$$e_r = \hat{d}_{post} - rank(pósteste)$$

A análise de variância é então realizada nesses escores de classificação de desvio residual.

6 Resultados

Para análise e resultados do banco de dados o software estatístico *Statistical Analysis System – SAS* foi utilizado e toda a programação estará disponível em anexo neste trabalho. Inicialmente será feita uma análise descritiva dos dados e, após análise, um modelo de análise de covariância será ajustado.

6.1 Análise descritiva

Além dos itens correspondentes a hábitos e conhecimentos alimentares, o questionário contem itens referentes a sexo, escolaridade dos responsáveis e idade dos respondentes. É possível notar pela Figura 2 que 52,93% dos alunos que participaram da pesquisa são do sexo masculino e 47,07% são do sexo feminino.

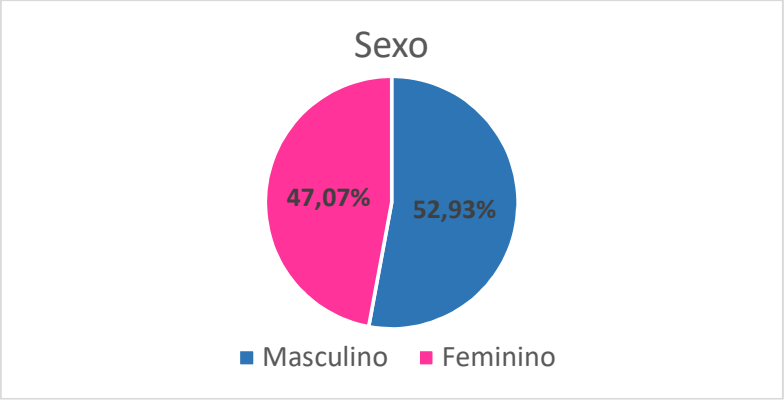

Figura 1 – Distribuição da variável sexo

A Figura 3 apresenta a distribuição das idades dos respondentes. Pode-se notar que há uma maior concentração de participantes entre 14 e 15 anos de idade, representando 71,05% do total. A distribuição da escolaridade dos responsáveis pelos respondentes é apresentado nas Figuras 4 e 5. Aproximadamente 16.05% das mães têm ensino superior completo enquanto 21,91% completaram o Ensino Médio, porém, 19.52% dos respondentes não sabiam o nível de escolaridade da sua mãe. Já 12,85% dos pais dos alunos têm ensino superior mas, 34,86% não sabem do seu nível de escolaridade.

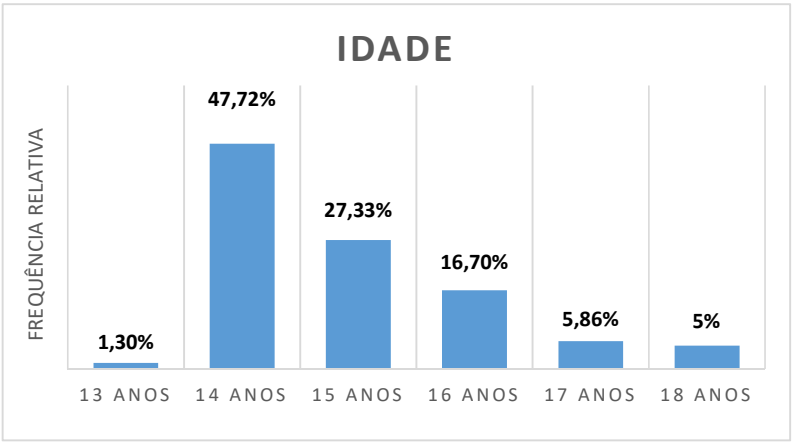

Figura 2 – Distribuição de idade

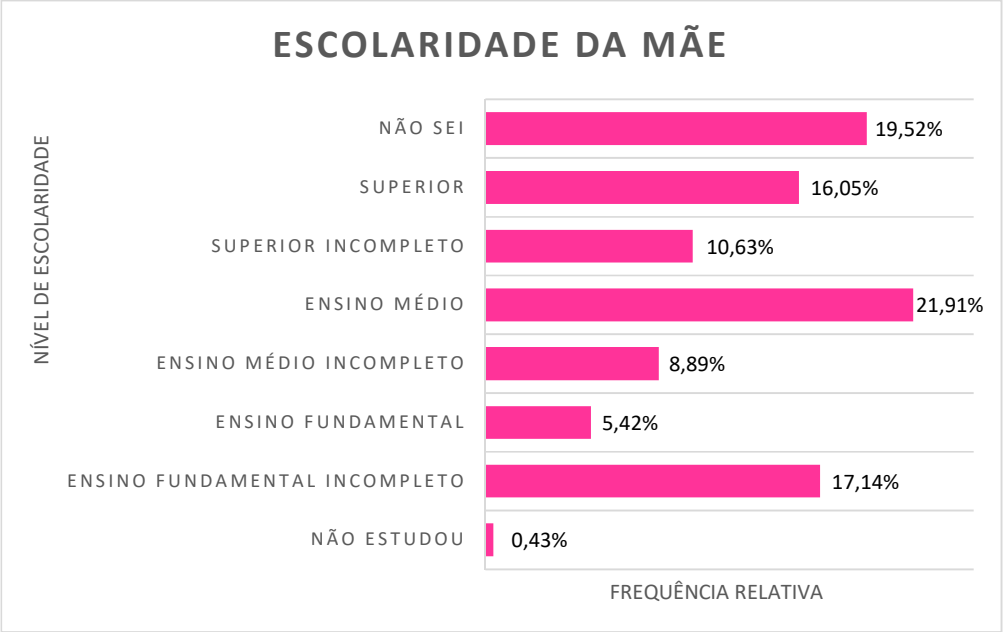

Figura 3 – Distribuição da escolaridade das mães dos respondentes

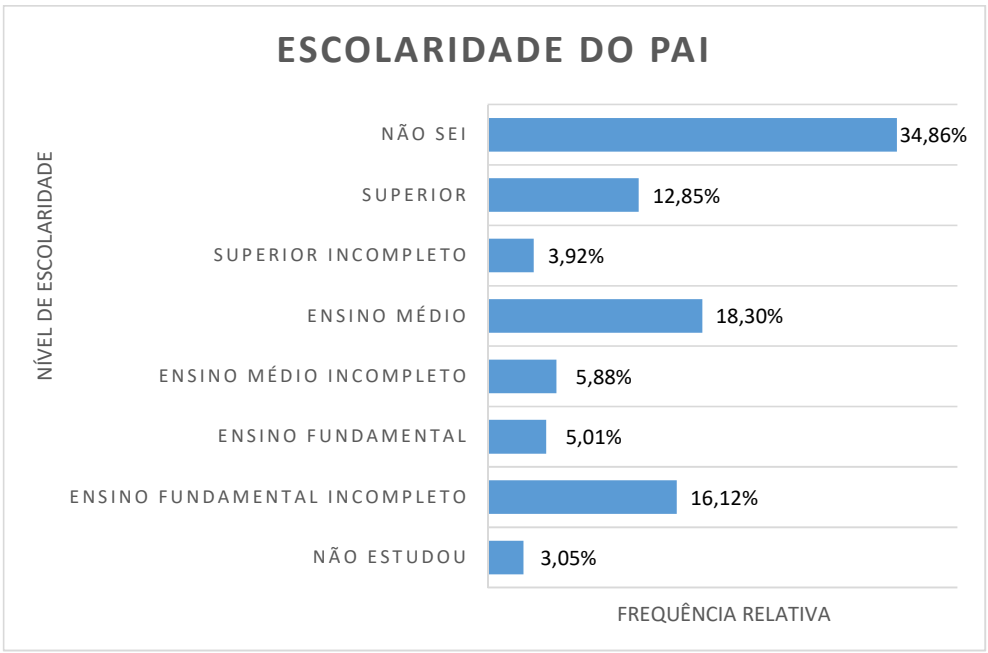

Figura 4 – Distribuição da escolaridade dos pais dos respondentes

Não se faz necessário o estudo das 88 variáveis para atingirmos nosso resultado. Por isso descrevemos abaixo as variáveis de interesse e que foram devidamente analisadas. A tabela 2 contém os nomes das variáveis, de acordo com a questão no relatório e suas respectivas descrições.

	Descrição
Q15	Ontem, quantas vezes comeu frutas frescas.
Q29	Costuma alimentar-se com a mãe, pai ou responsável.
Q30	Costuma comer assistindo à televisão ou estudando.
Q31	Costuma tomar o café da manhã.
Q32	Consideração (nota) sobre alimentação saudável.
Q33	Ciência (nota) sobre alimentação saudável.
Nfin1	Somatório das questões do tipo Verdadeiro ou Falso referentes a conhecimento.
Pint1	Presença nas 3 intervenções.

Tabela 2 – Variáveis estudadas e suas abreviações

As variáveis mencionadas após à reaplicação do questionário, terão a mesma descrição adicionadas a palavra "dois" para a devida abreviação: exemplo: Q29dois.

A tabela a seguir nos mostra a média e o intervalo de confiança, à 95%, da variável "Somatório das questões do tipo Verdadeiro ou Falso referentes a conhecimento", antes e após o tratamento, referentes ao conhecimento sobre uma alimentação saudável e adequada. Com esta análise inicial dos dados, foi observado que houve uma mudança no escore médio do conhecimento dos alunos, avaliando-se de 0 a 13, o conhecimento médio dos alunos no grupo de tratamento era de 7,54 e se deslocou para 9,26. Já o grupo controle estagnou-se em 7,19 em média.

Grupo	Frequência	Variável	Média	LI (IC 95%)	LS (IC 95%)
Tratamento	273	Nfin 1	7,54	7,20	7,88
		Nfin 2	9,26	8,91	9,61
Controle	188	Nfin 1	7,19	6,78	7,59
		Nfin 2	7,19	6,77	7,61

* LI – Limite inferior, LS – Limite superior e IC – Intervalo de confiança

* Nfin - Somatório das questões do tipo Verdadeiro ou Falso referentes a conhecimento

Tabela 3 – Médias e intervalos de confiança dos grupos

Começando a interpretação dos resultados pela análise descritiva das variáveis de interesse, comparadas aos dados após a reaplicação do questionário, o procedimento *proc freq* foi utilizado no *SAS*, pois, pelos dados da amostra ele estima a frequência populacional (Fq.), acumulada (F. acum.) e relativa (F.rel %) que auxiliam a compreensão.

Tabela 4 – Frequência da variável “Ontem, quantas vezes comeu frutas frescas” (Q15)

Ontem, quantas vezes comeu frutas frescas (Q15)				
Q15	Fq.	F.rel. %	F. acum,	F,rel acum. %
Não comeu	163	35,51	163	35,51
Comeu uma vez	129	28,10	292	63,62
Comeu duas vezes	99	21,57	391	85,19
Comeu três vezes	68	14,81	459	100,00

Ontem, quantas vezes comeu frutas frescas (Q15dois)				
Q15dois	Fq.	F.rel. %	F. acum	F,rel acum. %
Não comeu	179	38,9	179	38,9
Comeu uma vez	110	23,9	289	62,8
Comeu duas vezes	96	20,8	385	83,7
Comeu três vezes	75	16,30	460	100,00

Tabela 5 – Frequência da variável “Costuma alimentar-se com a mãe, pai ou responsável” (Q29)

Costuma alimentar-se com a mãe, pai ou responsável (Q29)				
Q29	Fq.	F.rel. %	F. acum,	F,rel acum. %
Não	71	16,06	71	16,06
Todos os dias	222	50,23	293	66,29
5 a 6 dias por semana	23	5,20	316	71,49
3 a 4 dias por semana	42	9,50	358	81,00
1 a 2 dias por semana	34	7,69	392	88,69
Sim, mas raramente	50	11,31	442	100,00

Costuma alimentar-se com a mãe, pai ou responsável (Q29dois)				
Q29dois	Fq.	F.rel. %	F. acum,	F,rel acum. %
Não	85	19,06	85	19,06
Todos os dias	208	46,64	293	65,70
5 a 6 dias por semana	34	7,62	327	73,32
3 a 4 dias por semana	35	7,85	362	81,17
1 a 2 dias por semana	44	9,87	406	91,03
Sim, mas raramente	40	8,97	446	100,00

Tabela 6 – Frequência da variável "Costuma comer assistindo à televisão ou estudando" (Q30)

Costuma comer assistindo à televisão ou estudando (Q30)				
Q30	Fq.	F.rel. %	F. acum,	F,rel acum. %
Não	81	18,33	81	18,33
Todos os dias	204	46,15	285	64,48
5 a 6 dias por semana	29	6,56	314	71,04
3 a 4 dias por semana	28	6,33	342	77,38
1 a 2 dias por semana	22	4,98	364	82,35
Sim, mas raramente	78	17,65	442	100,00

Costuma comer assistindo à televisão ou estudando (Q30dois)				
Q30dois	Fq.	F.rel. %	F. acum,	F,rel acum. %
Não	88	19,73	88	19,73
Todos os dias	197	44,17	285	63,90
5 a 6 dias por semana	31	6,95	316	70,85
3 a 4 dias por semana	35	7,85	351	78,70
1 a 2 dias por semana	30	6,73	381	85,43
Sim, mas raramente	65	14,57	446	100,00

Tabela 7 – Frequência da variável "Costuma tomar o café da manhã" (Q31)

Costuma tomar o café da manhã (Q31)				
Q31	Fq.	F.rel. %	F. acum,	F,rel acum. %
Não	107	24,21	107	24,21
Todos os dias	221	50,00	328	74,21
5 a 6 dias por semana	16	3,62	344	77,83
3 a 4 dias por semana	33	7,47	377	85,29
1 a 2 dias por semana	16	3,62	393	88,91
Sim, mas raramente	49	11,09	442	100,00

Costuma tomar o café da manhã (Q31dois)				
Q31dois	Fq.	F.rel. %	F. acum,	F,rel acum. %
Não	97	21,85	97	21,85
Todos os dias	221	49,77	318	71,62
5 a 6 dias por semana	20	4,50	338	76,13
3 a 4 dias por semana	31	6,98	369	83,11
1 a 2 dias por semana	25	5,63	394	88,74
Sim, mas raramente	50	11,26	444	100,00

Tabela 8 – Frequência da variável "Consideração (nota) sobre alimentação saudável" (Q32)

Consideração (nota) sobre alimentação saudável (Q32)				
Q32	Fq.	F.rel. %	F. acum,	F,rel acum. %
0	13	2,95	13	2,95
0,5	1	0,23	14	3,18
1	4	0,91	18	4,09
2	9	2,05	27	6,14
3	21	4,77	48	10,91
4	29	6,59	77	17,50
5	103	23,41	180	40,91
5,8	1	0,23	181	41,14
6	70	15,91	251	57,05
6,5	6	1,36	257	58,41
7	98	22,27	355	80,68
7,5	4	0,91	359	81,59
8	53	12,05	412	93,64
8,5	4	0,91	416	94,55
9	13	2,95	429	97,50
9,5	3	0,68	432	98,18
10	8	1,92	440	100,00

Consideração (nota) sobre alimentação saudável (Q32dois)				
Q32dois	Fq.	F.rel. %	F. acum,	F,rel acum. %
0	13	2,92	13	2,92
0,2	1	0.22	14	3,15
1	5	1,12	19	4,27
2	6	1,35	25	5,62
3	11	2,47	36	8,09
4	29	6,52	65	14,61
4,5	2	0.45	67	15,06
5	88	19,78	155	34,83
5,5	3	0,67	158	35,51
6	72	16,18	230	51,69
6,5	1	0,22	231	51,91
7	96	21,57	327	73,48
7,5	7	1,57	334	75,06
8	74	16,63	408	91,69
8,5	5	1,12	413	92,81
9	20	4,49	433	97,30
10	12	2,70	445	100,00

Tabela 9 – Frequência da variável "Ciência (nota) sobre alimentação saudável" (Q33)

Ciência (nota) sobre alimentação saudável (Q33)				
Q33	Fq.	F.rel. %	F. acum,	F,rel acum. %
0	18	4,09	18	4,09
0,7	1	0,23	19	4,32
1	3	0,68	22	5,00
2	9	2,05	31	7,05
3	13	2,95	44	10,00
4	17	3,86	61	13,86
4,5	1	0,23	62	14,09
5	92	20,91	154	35,00
6	48	10,91	202	45,91
6,5	2	0,45	204	46,36
7	78	17,73	282	64,09
7,5	3	0,68	285	64,77
8	64	14,55	349	79,32
8,5	1	0,23	350	79,55
9	45	10,23	395	89,77
9,5	1	0,23	396	90,00
10	44	10,00	440	100,00

Ciência (nota) sobre alimentação saudável (Q33dois)				
Q33dois	Fq.	F.rel. %	F. acum,	F,rel acum. %
0	14	3,14	14	3,14
0,5	1	0,22	15	3,36
1	3	0,67	18	4,04
2	5	1,12	23	5,16
3	9	2,02	32	7,17
4	20	4,48	52	11,66
4,5	1	0,22	53	11,88
5	80	17,94	133	29,82
6	53	11,88	186	41,70
7	80	17,94	266	59,64
7,5	6	1,35	272	60,99
7,8	1	0,22	273	61,21
8	78	17,49	351	78,70
8,5	3	0,67	354	79,37
8,9	1	0,22	355	79,60
9	59	11,21	405	90,81
9,5	3	0,67	408	91,48
9,6	1	0,22	409	91,70
10	37	8,30	446	100,00

Percebe-se a priori pequenas mudanças nas frequências das variáveis estudas, porém, não há um informação tão relevante para essa análise inicial. Informações expostas mais a critério exploratório e ilustrativo.

6.2 Análise de covariância

6.2.1 Paramétrica

- **Variável "Somatório das questões do tipo Verdadeiro ou Falso referentes a conhecimento" (Nfin)**

A covariável "Somatório das questões do tipo Verdadeiro ou Falso referentes a conhecimento" (Nfin1) relaciona-se significativamente com a variável resposta (p-valor < 0,0001), isto implica que ela realmente influencia na média desta variável, e a inclusão dela no modelo aumenta a precisão das estimativas. Podemos analisar o devido resultado pelo quadro da análise de covariância.

Fonte de variação	G.L.	SQ	QM	F	Pr > F
Pint1	1	41,2082	41,2082	7,14	0,0078
Grupo	1	69,3710	69,3710	12,02	0,0006
Nfin1	1	1278,2314	1278,2314	221,46	<0,0001

Tabela 10 – Quadro da análise de variância para ANCOVA

Devemos também testar a hipótese nula que os coeficientes de regressão da população dentro dos grupos são diferentes, isto é $\beta_1 \neq \beta_2 \neq \cdots \neq \beta_w$, ou seja se as retas não são paralelas. Com este teste foi obtido o $F_{obs} = 102{,}47$ e o p-valor menor que 0,0001, considerando um nível de significância de 95%, este teste evidencia que o modelo assume que as inclinações relacionadas aos escores pós-tratamento aos escores de pré-tratamento são iguais para todos os tratamentos, além de que as regressões lineares simples tem inclinação diferente de zero ($\beta \neq 0$). Os resultados são apresentados na tabela seguinte.

Fonte de variação	G.L.	SQ	QM	F	P-valor
Modelo	3	1774,286	591,429	102,47	< 0,0001
Erro	457	2637,749	5,772		
Total corrigido	460	4412,035			

Tabela 11 – Comparando duas ou mais retas de regressão

Através da comparação exploratória das médias e do seu intervalo de confiança, vistas anteriormente (Tabela 3), para as variáveis Nfin1 e Nfin2, temos evidências para afirmar uma suposta distinção entre os tratamentos aplicados.

A tabela a seguir nos apresenta as médias de tratamentos ajustadas pela covariável, que é o estimador de mínimos quadrados de $\mu + \tau_i$ e dois importantes testes, testando as hipóteses da média da variável Nfin2 = 0 em ambos os grupos, e se existe igualdade entre as médias dos grupos.

Grupo	Nfin2 LSMEAN	Erro padrão	HO: LSMEAN = 0	H0: LSMEAN1 = LSMEAN 2
			Pr > |t|	Pr > |t|
Controle	7,7161	0.2309	< 0,0001	0,0006
Tratamento	8,8988	0,1785	< 0,0001	

* Nfin2 - Somatório das questões do tipo Verdadeiro ou Falso referentes a conhecimento pós tratamento

Tabela 12 – Teste de comparação de médias

Com p-valor apresentado igual a 0,0006, considerando um nível de significância de 95% o teste de Tukey-Kramer indica que há evidências que existe diferença entre as médias dos tratamentos. A tabela abaixo apresenta os pesos médios da variável Nfin2 baseada nos grupos controle e tratamento, com e sem o ajuste da covariável.

Grupo	Média	Média ajustada
Controle	7,19	7,7161
Tratamento	9,26	8,8988

Tabela 13 – Comparação das médias ajustadas pela covariável

Diferença	Intervalo de confiança
-1,1827	(- 1,8530 ; - 0,5123)

Tabela 14 – Diferença entre as médias ajustadas

No gráfico a seguir podemos observar como a média ajustada pelo modelo de covariância se desloca entre os grupos controle e tratamento, mostrando um aumento na média significativo.

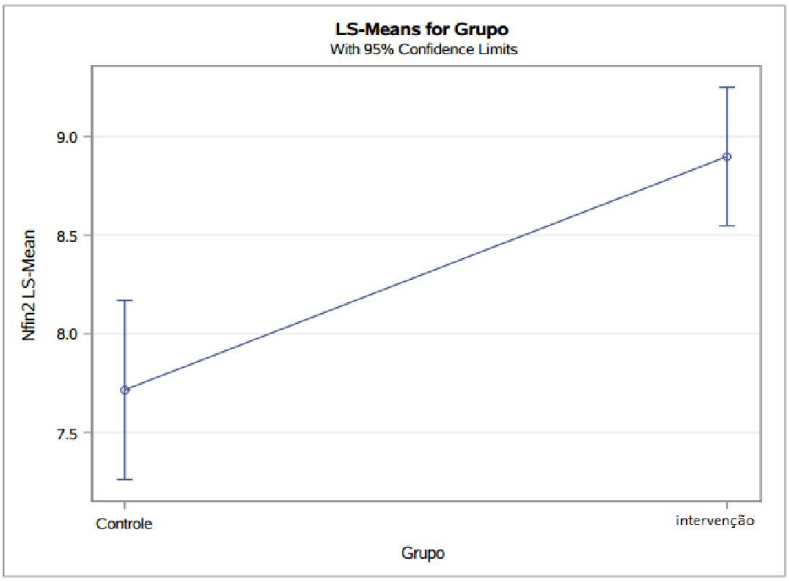

Figura 5 - Médias ajustadas da variável "Somatório das questões do tipo Verdadeiro ou Falso referentes a conhecimento"

Concluindo, pelo gráfico e pelas médias ajustadas pela análise de covariância, observa-se que houve um aumento significativo no nível médio de conhecimento dos alunos, sobre a importância de uma alimentação saudável. Interessante também é analisar a diferença do valor médio entre as variáveis Nfin2 e Nfin1 para os dois grupos. Para o grupo controle o valor $t_{obs} = 0{,}03$ e o p-valor = 0,9755, portanto temos evidências para não rejeitar a hipótese nula que afirma que não há diferença nas médias das variáveis, com 95% de significância. Já para o grupo de tratamento $t_{obs} = 9{,}84$ e p-valor < 0,0001, logo a diferença entre as médias das variáveis se faz evidente.

Por fim apresento a tabela das estimativas dos parâmetros para a devida adequação do modelo estimado.

Parâmetro	Estimativa	Erro padrão	t	Pr > $\|t\|$
Intercepto	4,1290	0,4280	9,65	<0.0001
Pint1	0,8997	0,3367	2,67	0,0078
Grupo controle	- 1,1827	0,3412	- 3,47	0,0006
Grupo tratamento	0,0000			
Nfin1	0,5907	0,0397	14,88	<0,0001

Tabela 15 – Tabela das estimativas dos parâmetros

Para fazer a verificação da adequabilidade para ANCOVA a análise de resíduos se mostra muito útil. Na aplicação em questão, devido a grandeza do tamanho da amostra, a identificação de alguma estrutura considerável nos gráficos de resíduos se torna rara, além disso algumas poucas observações dificilmente alterariam significativamente o ajuste do modelo.

Se as suposições são satisfeitas e o modelo está ajustado, os gráfico de resíduos versus valor predito e resíduos versus covariável não devem apresentar nenhuma estrutura, já o gráfico de probabilidade normal deve mostrar os pontos próximos a uma linha reta diagonal. Com o "delay" previsto e exposto acima, temos as seguintes figuras. Logo consideramos um bom ajuste do modelo na aplicação.

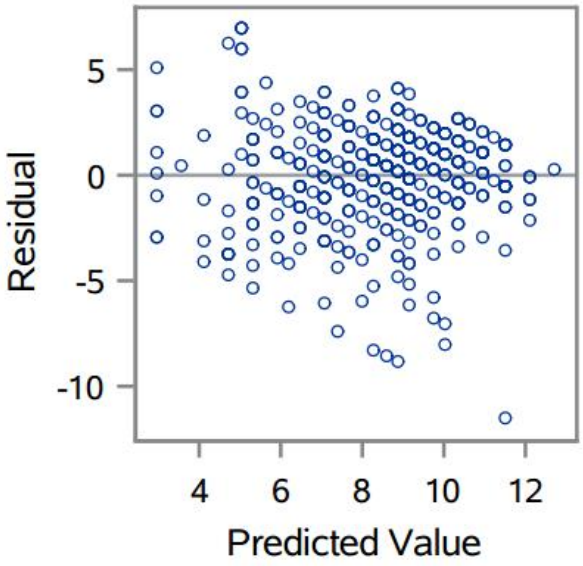

Figura 6 – Gráfico resíduos versus valor predito

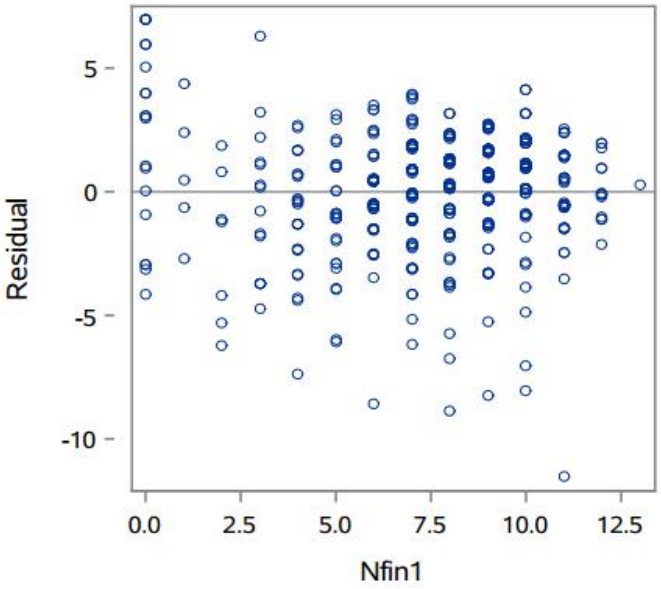

Figura 7 – Gráfico resíduos versus covariável

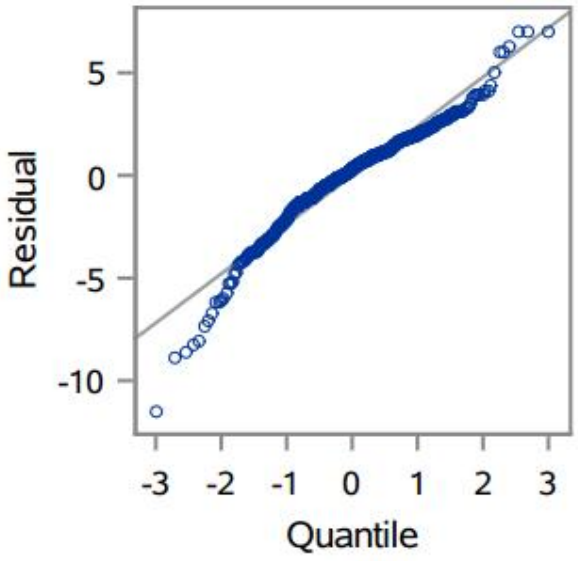

Figura 8 – Gráfico de probabilidade normal

- **Variável "Consideração (nota) sobre alimentação saudável" (Q32)**

A covariável "Consideração (nota) sobre alimentação saudável" (Q32) relaciona-se significativamente com a variável resposta (p-valor < 0,0001), seu $F_{obs} = 264{,}16$ encontra-se na região crítica do teste F-Snedecor e isto implica que ela realmente influencia na média desta variável, a inclusão dela no modelo se faz necessária, pois, aumenta a precisão das estimativas.

Testanto também a hipótese nula que os coeficientes de regressão da população dentro dos grupos são diferentes, temos que $F_{obs} = 89{,}22$ e p-valor < 0,0001, temos evidências a um nível de significância de 95%, que o modelo assume que as inclinações são iguais, gerando

retas paralelas para todas os tratamentos com $\beta \neq 0$. Os resultados são apresentados na tabela seguinte.

Fonte de variação	G.L.	SQ	QM	F	P-valor
Modelo	3	688,8420	229,6140	89,22	< 0,0001
Erro	424	1091,1852	2,5735		
Total corrigido	427	1780,0273			

Tabela 16 – Comparando duas ou mais retas de regressão

Através da comparação exploratória das médias e do seu intervalo de confiança expostas na tabela 27, para as variáveis Q32 e Q32dois, percebemos um pequeno aumento na média após o tratamento para ambos os grupos.

Grupo	Frequência	Variável	Média	LI (IC 95%)	LS (IC 95%)
Tratamento	273	Q32	5,98	5,76	6,21
		Q32dois	6,21	5,98	6,44
Controle	188	Q32	5,70	5,37	6,02
		Q32dois	6,02	5,70	6,35

* Q32 - Consideração (nota) sobre alimentação saudável pré tratamento

* Q32dois - Consideração (nota) sobre alimentação saudável pós tratamento

Tabela 17 – Médias e intervalos de confiança dos grupos

A tabela de resultados a seguir nos apresenta dois testes para as médias de tratamentos ajustadas pela covariável, testando as hipóteses da média da variável Q32dois = 0 em ambos os grupos, e se existe igualdade entre as médias dos grupos.

			HO: LSMEAN = 0	H0: LSMEAN1 = LSMEAN 2
Grupo	Q32dois LSMEAN	Erro padrão	Pr > $\|t\|$	Pr > $\|t\|$
Controle	6,1605	0.1594	< 0,0001	0,7948
Tratamento	6,0982	0,1266	< 0,0001	

* Q32dois - Consideração (nota) sobre alimentação saudável pós tratamento

Tabela 18 – Teste de comparação de médias

Com p-valor apresentado igual a 0,79486, considerando um nível de significância de 95% o teste de Tukey-Kramer evidencia que não existe diferença entre as médias dos grupos tratamento e controle. A tabela abaixo apresenta a diferença entre os pesos médios da variável Q32dois ajustada pela covariável.

Diferença	Intervalo de confiança
0,0623	(- 0,4081 ; 0,5327)

Tabela 19 – Diferença entre as médias ajustadas

Pelo gráfico podemos observar como a média ajustada pelo modelo do covariância se desloca entre os grupos controle e tratamento, mostrando uma pequena diminuição na média, portanto, não significativo.

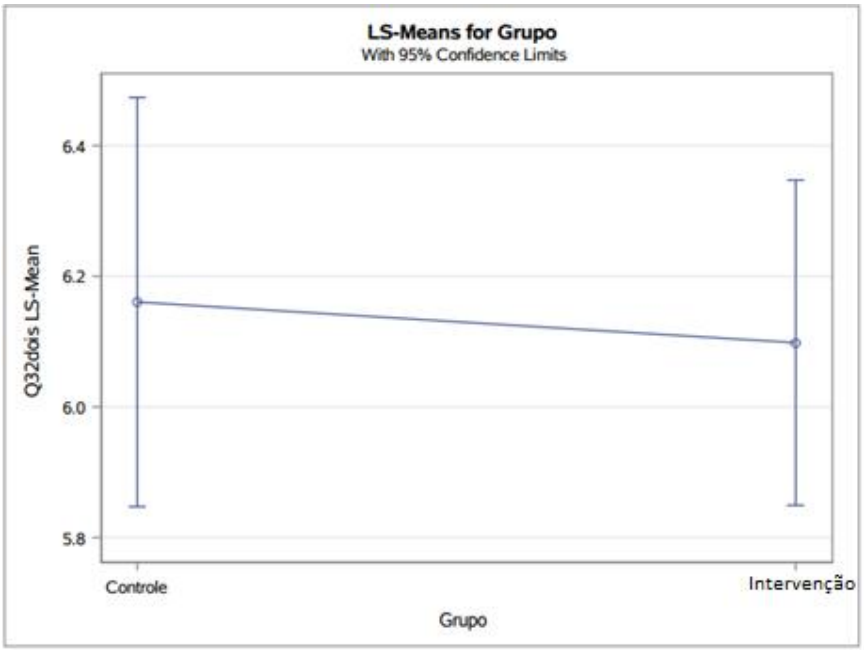

Figura 9 - Médias ajustadas da variável Consideração (nota) sobre alimentação saudável

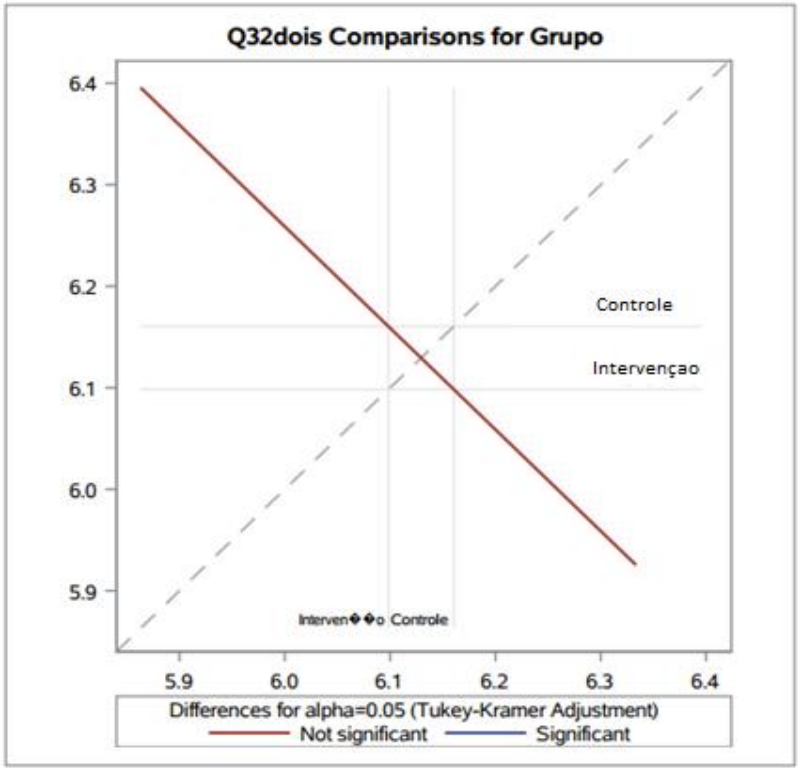

Figura 10 - Ajuste para múltiplas comparações das médias por Tukey-Krame

Com auxílio do gráfico anterior e pelas estatísticas citadas, observa-se que houve uma diminuição média na consideração dos respondentes sobre sua alimentação saudável, porém não significativa para o estudo.

Para a análise da diferença do valor médio entre as variáveis Q32dois e Q32 (Consideração (nota) sobre alimentação saudável pós e pré tratamento, respectivamente) para os dois grupos, obtemos para o grupo controle o p-valor = 0,0215 e para o grupo de tratamento o p-valor = 0,0355 portanto temos evidências para rejeitar a hipótese nula que afirma que não há diferença nas médias das variáveis, com 95% de significância.

Como verificação da adequabilidade do modelo, fazemos o diagnóstico pela análise de resíduos.

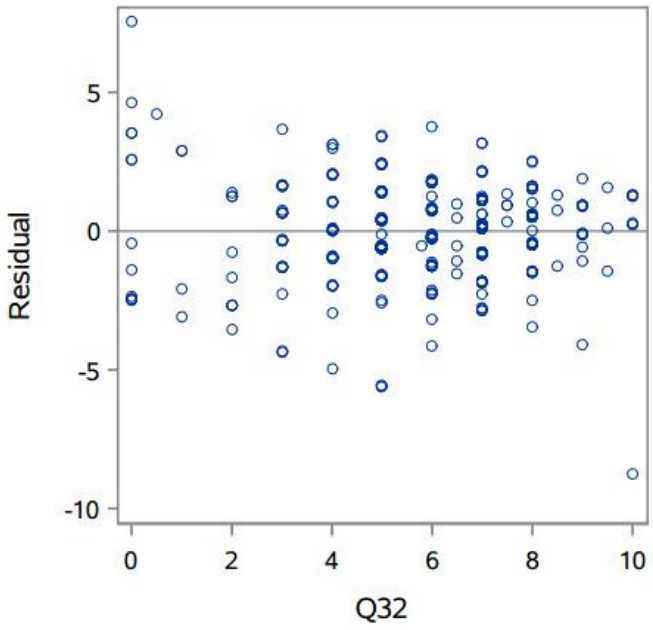

Figura 11 – Gráfico resíduos versus covariável

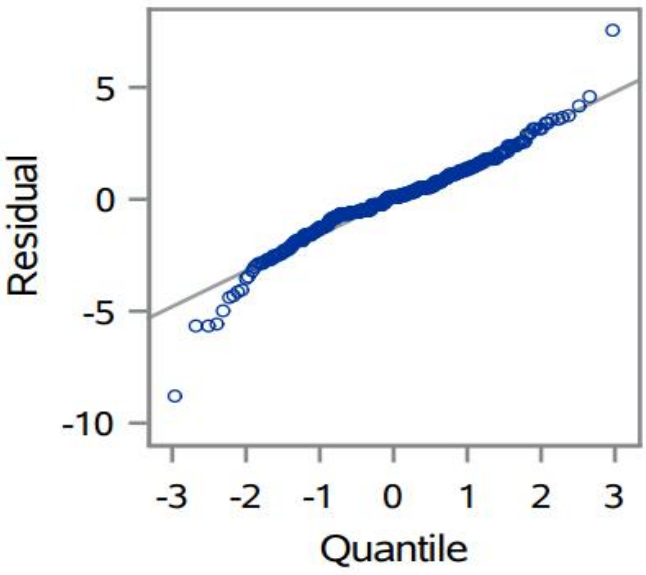

Figura 12 – Gráfico de probabilidade normal

- **Variável "Ciência (nota) sobre alimentação saudável" (Q33)**

Assim como as covariáveis "Somatório das questões do tipo Verdadeiro ou Falso referentes a conhecimento" (Nfin) e "Consideração (nota) sobre alimentação saudável" (Q32), a variável "Ciência (nota) sobre alimentação saudável" (Q33) também tem uma relação significativa com a variável resposta por apresentar um $F_{obs} = 169{,}83$ e um p-valor < 0,0001 aferidos na função "proc glm" do SAS. Isto implica que ela realmente influencia na média da variável resposta, logo, a inclusão dela no modelo se faz necessária.

Testanto também a hipótese nula que os coeficientes de regressão dentro dos grupos são diferentes, temos que $F_{obs} = 61{,}90$ e p-valor < 0,0001, a um nível de significância de 95%, temos evidências para assumir que as inclinações são iguais e as retas paralelas para todas os tratamentos com $\beta \neq 0$ Os resultados são apresentados na tabela seguinte.

Fonte de variação	G.L.	SQ	QM	F	P-valor
Modelo	3	654,8678	218,2893	61,90	< 0,0001
Erro	424	1495,1327	3,5262		
Total corrigido	427	2150,0005			

Tabela 20 – Comparando duas ou mais retas de regressão

Através da comparação exploratória das médias expostas na tabela 21, para as variáveis Q33 e Q33dois (Ciência (nota) sobre alimentação saudável pré e pós tratamento, respectivamente), percebemos um pequeno aumento na média após o tratamento para ambos os grupos.

Grupo	Frequência	Variável	Média	LI (IC 95%)	LS (IC 95%)
Tratamento	273	Q32	6,47	6,20	6,75
		Q32dois	6,95	6,70	7,20
Controle	188	Q32	6,41	6,03	6,80
		Q32dois	6,26	5,90	6,62

Tabela 21 – Médias e intervalos de confiança dos grupos

A tabela de resultados a seguir nos apresenta dois testes para as médias de tratamentos ajustadas pela covariável, testando as hipóteses da média da variável Q33dois = 0 em ambos os grupos, e se existe igualdade entre as médias dos grupos.

			HO: LSMEAN = 0	H0: LSMEAN1 = LSMEAN 2
Grupo	Q33dois LSMEAN	Erro padrão	Pr > $\|t\|$	Pr > $\|t\|$
Controle	6,4507	0.1868	< 0,0001	0,1990
Tratamento	6,8117	0,1484	< 0,0001	

* Q33dois - Ciência (nota) sobre alimentação saudável pós tratamento

Tabela 22 – Teste de comparação de médias

Com p-valor apresentado igual a 0,1990, considerando um nível de significância de 95% o teste de Tukey-Kramer evidencia que não existe diferença entre as médias dos grupos tratamento e controle. A tabela abaixo apresenta a diferença entre médias da variável Q32dois ajustada pela covariável.

Diferença	Intervalo de confiança
- 0,3610	(- 0,9127 ; 0,1906)

Tabela 23 – Diferença entre as médias ajustadas

No gráfico a seguir podemos observar como a média ajustada pelo modelo do covariância se desloca entre os grupos controle e tratamento, mostrando um aumento na média, porém, não significativo.

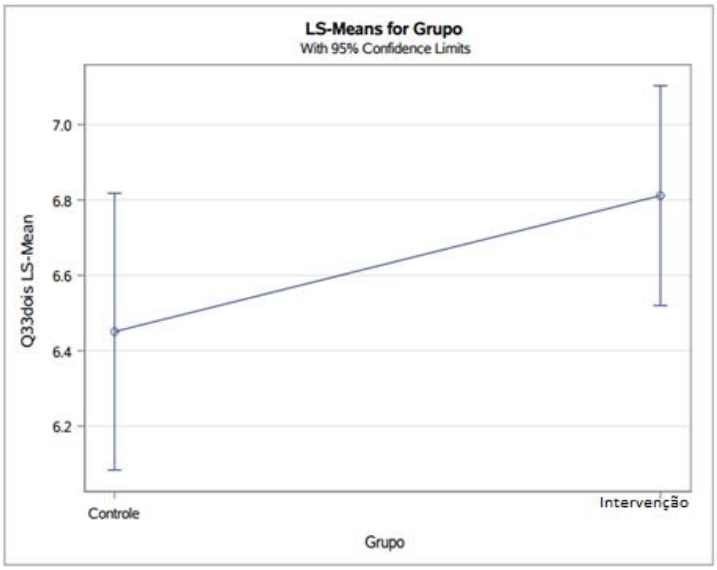

Figura 13 - Médias ajustadas da variável Q32dois

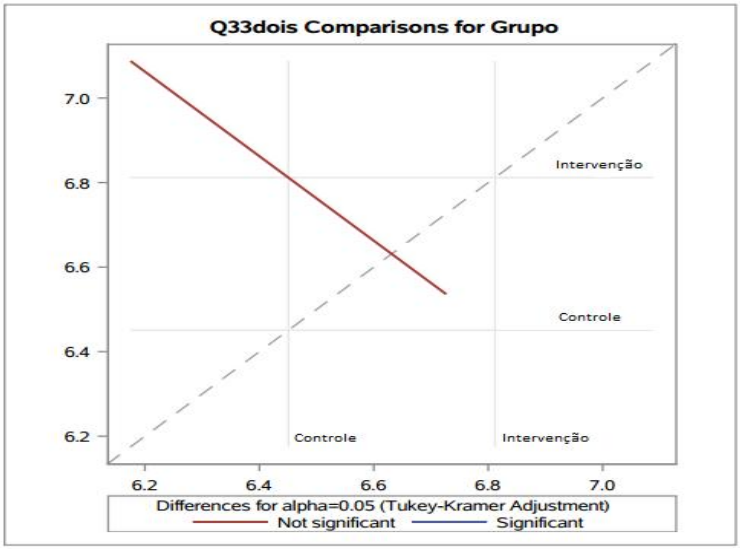

Figura 14 - Ajuste para múltiplas comparações das médias por Tukey-Krame

Pelos gráficos expostos e pelas médias ajustadas pela análise de covariância, observa-se que houve um aumento, porém, não significativo no nível médio de auto avaliação de conhecimentos dos alunos, sobre a importância de uma alimentação saudável. Em contrapartida, em resultados anteriores (estudo da variável "Somatório das questões do tipo Verdadeiro ou Falso referentes a conhecimento"), já concluímos que há evidências de que o nível médio de conhecimento dos respondentes aumentou. Características próprias de questões um tanto subjetivas, sem uma referência real de comparação.

Para a análise da diferença do valor médio entre as variáveis Q32dois e Q32 para os dois grupos, obtemos para o grupo controle o p-valor = 0,4381, portanto temos evidências para não rejeitar a hipótese nula que afirma que não há diferença nas médias das variáveis, com 95% de significância. Já para o grupo de tratamento o p-valor = 0,0061 portanto a diferença entre as médias das variáveis se faz evidente.

Como diagnóstico do modelo temos os gráficos das figuras 16 e 17 como auxiliadores para análise.

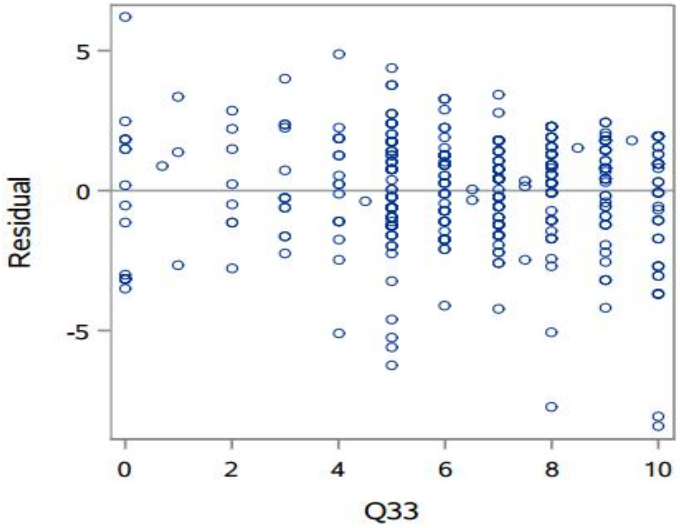

Figura 15 – Gráfico resíduos versus covariável

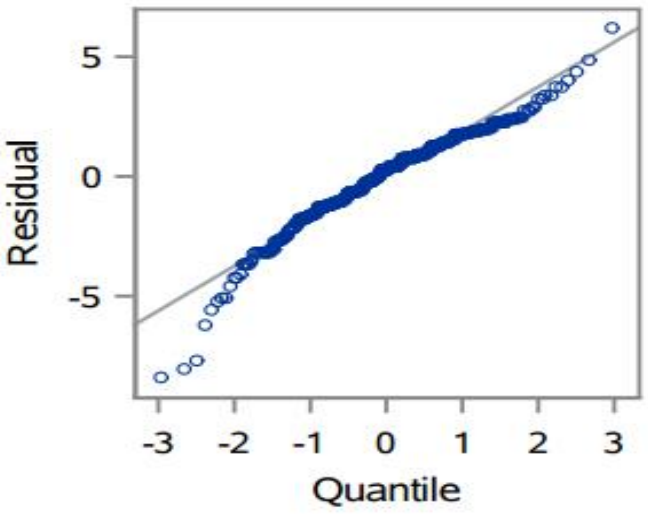

Figura 16 – Gráfico de probabilidade normal

6.2.2 Análise de covariância não paramétrica

Ao contrário da forma da análise dos resultados para as variáveis de interesse paramétricas, compilaremos as variáveis estudas “Ontem, quantas vezes comeu frutas frescas” (Q15), “Costuma alimentar-se com a mãe, pai ou responsável” (Q29), “Costuma comer assistindo à televisão ou estudando” (Q30) e “Costuma tomar o café da manhã” (Q31), por apresentarem resultados, formas de tratamento e aplicação similares.

Devido à violação nas suposições do modelo paramétrico, a análise de covariância não paramétrica se faz necessária. Como visto, que “rank ANCOVA”, geralmente é mais poderoso que o ANCOVA paramétrico, além de estar de acordo com estudos de simulação de Conover e Iman (1982), que demonstraram que o teste de Quade e a ANCOVA paramétrica nos postos geralmente estão em acordo virtual, é viável que tomemos essas técnicas como referência.

Para as quatro variáveis em questão, primeiramente, transformamos os escores do pré-teste e do pós-teste, classificando-os por ranking. A única que apresenta uma pequena variação na forma de resposta, é a variável “Ontem, quantas vezes comeu frutas frescas” (Q15), as respostas são do tipo: não comi frutas frescas; comi uma vez; comi duas vezes; comi três vezes no dia anterior. Naturalmente existe uma ordem entre elas, por isso a classificamos como uma variável categórica ordinal.

Já as demais variáveis apresentam respostas similares do tipo: não; sim, todos os dias; sim, 5 a 6 dias por semana; sim, 3 a 4 dias por semana; sim, 1 a 2 dias por semana; sim, mas apenas raramente. Nas tabelas a seguir encontram-se as posições e os escores estabelecidos por ranking encontradas para cada variável pré e pós teste.

Q15		Q15dois	
Rank		Rank	
0	81,5	0	89,5
1	227,0	1	233,0
2	341,0	2	335,5
3	424,5	3	421,0

*Q15 - Ontem, quantas vezes comeu frutas frescas

*Q15dois - Ontem, quantas vezes comeu frutas frescas (pós tratamento)

Tabela 24 – Ranks e intervalos da variável Q15

Q29		Q29dois	
Rank		Rank	
0	35,0	0	42,0
1	93,0	1	103,5
2	132,0	2	145,0
3	168,0	3	183,0
4	199,5	4	215,5
5	320,5	5	331,0

*Q29 - Costuma alimentar-se com a mãe, pai ou responsável

*Q29dois - Costuma alimentar-se com a mãe, pai ou responsável (pós tratamento)

Tabela 25 – Ranks e intervalos da variável Q29

Q30		Q30dois	
Rank		Rank	
0	40,5	0	43,0
1	119,0	1	117,0
2	168,5	2	163,0
3	193,5	3	193,0
4	221,5	4	223,5
5	333,0	5	334,5

*Q30 - Costuma comer assistindo à televisão ou estudando

*Q30dois - Costuma comer assistindo à televisão ou estudando (pós tratamento)

Tabela 26 – Ranks e intervalos da variável Q30

Q31		Q31dois	
Rank		Rank	
0	53,0	0	49,0
1	128,5	1	121,0
2	159,5	2	156,0
3	183,0	3	183,0
4	206,5	4	208,5
5	421,5	5	323

*Q30 - Costuma tomar o café da manhã

*Q30dois - Costuma tomar o café da manhã (pós tratamento)

Tabela 27 – Ranks e intervalos da variável Q31

As classificações das pontuações foram então convertidas em pontuações de desvio de grau médio, subtraindo a classificação média de cada observação, formando um escore de classificação do desvio residual. Calcula-se um regressão linear nessas classificações, depois sujeitamos os resíduos a ANOVA,

Para as quatro variáveis, o teste de Quade apresentou um resultado não significativo para a ANOVA dos resíduos. O p-valor encontrado foi de p = 0,3588; p = 0,4099; p = 0,2366; p = 0,7188; respectivamente, para Q15, Q29, Q30 e Q31. Por este teste, temos evidências à 95% de significância, para não rejeitarmos a hipótese nula que os valores médios dos tratamentos são iguais.

Segue abaixo os gráficos da distribuição residual por grupo com a estatística do teste F e o p-valor anexado para cada variável.

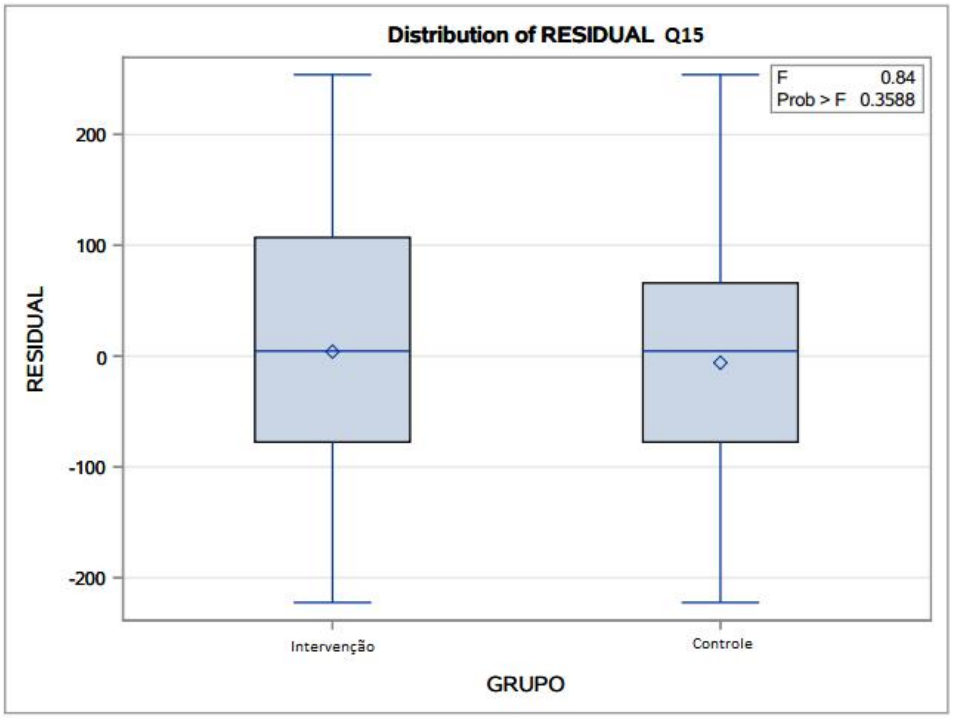

Figura 17 – Boxplot da distribuição residual “Ontem, quantas vezes comeu frutas frescas”

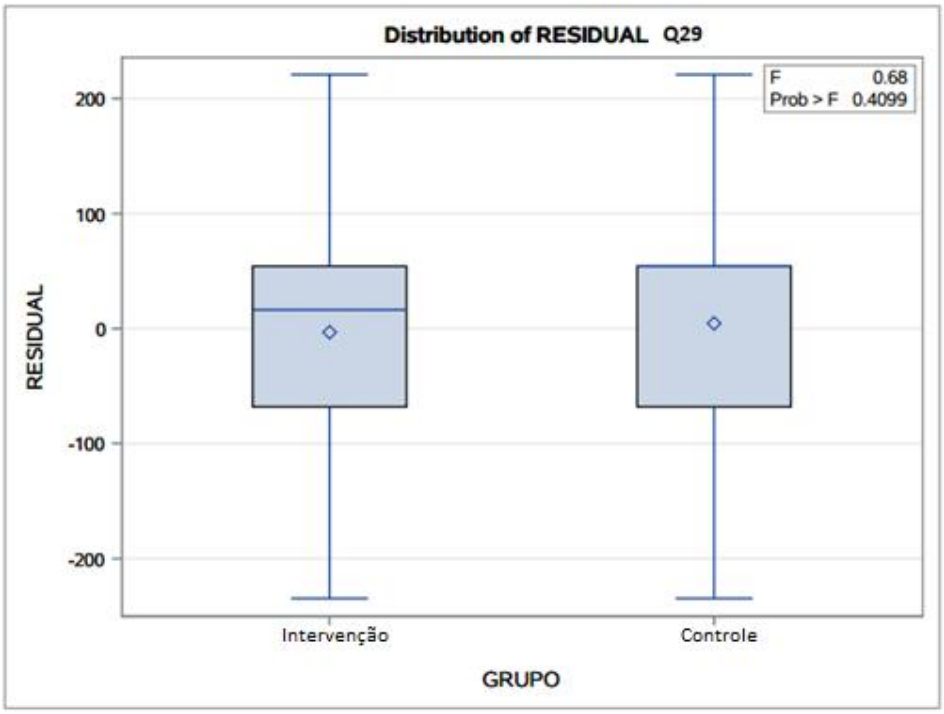

Figura 18 – Boxplot da distribuição residual “Costuma alimentar-se com a mãe, pai ou responsável”

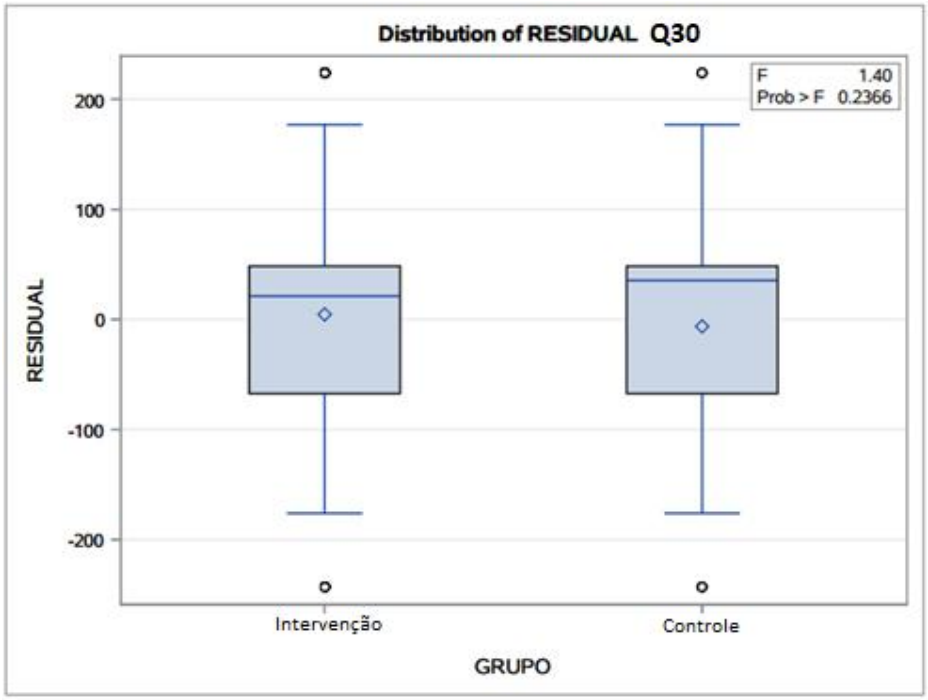

Figura 19 – Boxplot da distribuição residual “Costuma comer assistindo à televisão ou estudando”

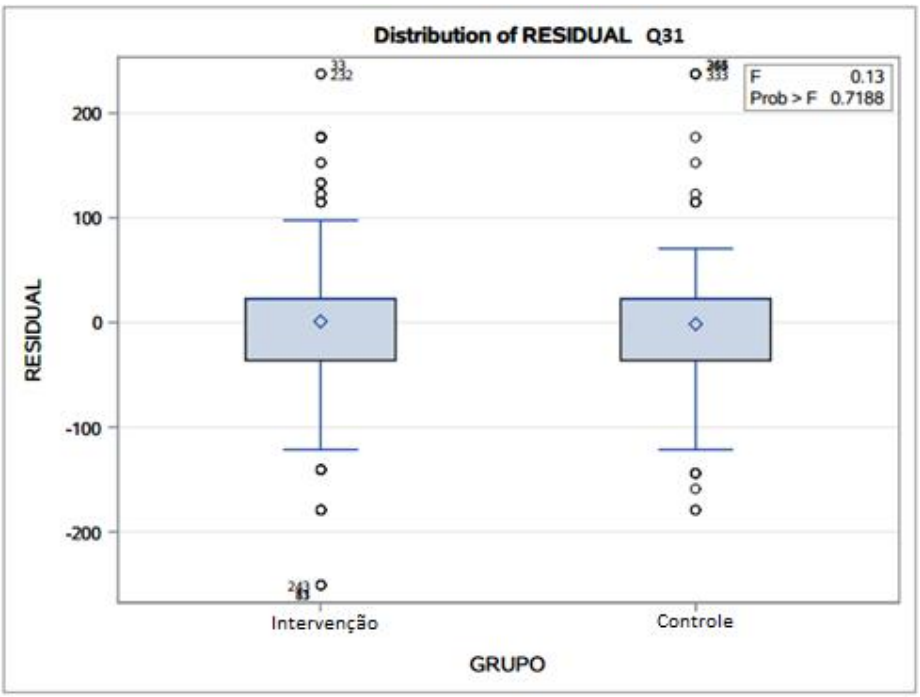

Figura 20 – Boxplot da distribuição residual “Costuma tomar o café da manhã”

Concluímos com o auxílio dos gráficos e das estatísticas que, pelo teste de Quade, temos evidências, a um nível de 95% de significância, para assumirmos que a média dos dois grupos de tratamento são iguais. Aceitamos portanto a hipótese nula. É esperado que o teste de análise de covariância paramétrico nos ranks, aponte os mesmos resultados.

Já para análise paramétrica nos ranks, as quatro covariáveis (rankpre) apontaram uma relação significativa com o variável resposta, influenciando, portanto, na média desta variável. As estatísticas F observadas, foram todas maiores que 50 e consequentemente pertenciam à região crítica, logo, o p-valor foi menor que 0,0001.

O teste da hipótese nula que os coeficientes de regressão dentro dos grupos são diferentes ($\beta_1 \neq \beta_2 \neq \cdots \neq \beta_w$), ou seja, que as retas não são paralelas, também deram significativos para as variáveis em estudo. Obtemos, $F_{obs} = 50{,}21$; $F_{obs} = 105{,}03$; $F_{obs} = 120{,}57$ e $F_{obs} = 368{,}64$ respectivamente para as variáveis Q15, Q29, Q30 e Q31 e p-valor menor que 0,0001. Considerando um nível de significância de 95%, este teste evidencia que o modelo assume que as inclinações relacionadas são iguais para todas os tratamentos, além de que as regressões lineares simples tem inclinação diferente de zero ($\beta \neq 0$).

As tabelas 28, 29, 30 e 31 a seguir nos apresenta os resultados para o teste de Tukey, além de apresentar o intervalo de confiança e o erro padrão da média.

Grupo	Rankpost LSMEAN	Intervalo de confiança		Erro padrão	HO: LSMEAN = 0	H0: LSMEAN1 = LSMEAN 2
					Pr > $\|t\|$	Pr > $\|t\|$
Controle	223,5419	206,995	240,088	8.4195	< 0,0001	0,3595
Tratamento	233,5742	219,893	247,255	6,9618	< 0,0001	

Tabela 28 – Comparação de médias e intervalo de confiança – VARIÁVEL Q15 – "Ontem, quantas vezes comeu frutas frescas"

Grupo	Rankpost LSMEAN	Intervalo de confiança		Erro padrão	HO: LSMEAN = 0	H0: LSMEAN1 = LSMEAN 2
					Pr > $\|t\|$	Pr > $\|t\|$
Controle	220,0106	205,905	234,116	7.1764	< 0,0001	0,4103
Tratamento	212,2523	200,284	224,220	6,0891	< 0,0001	

Tabela 29 – Comparação de médias e intervalo de confiança – VARIÁVEL Q29 – "Costuma alimentar-se com a mãe, pai ou responsável"

Grupo	Rankpost LSMEAN	Intervalo de confiança		Erro padrão	HO: LSMEAN = 0 Pr > |t|	H0: LSMEAN1 = LSMEAN 2 Pr > |t|
Controle	209,0983	195,253	240,943	7.0438	< 0,0001	0,2342
Tratamento	220,1534	208,361	231,945	5,9958	< 0,0001	

Tabela 30 – Comparação de médias e intervalo de confiança – VARIÁVEL Q30 – "Costuma comer assistindo à televisão ou estudando"

Grupo	Rankpost LSMEAN	Intervalo de confiança		Erro padrão	HO: LSMEAN = 0 Pr > |t|	H0: LSMEAN1 = LSMEAN 2 Pr > |t|
Controle	213,0679	202,814	223,332	5.2168	< 0,0001	0,7191
Tratamento	215,5294	206,835	224,223	4,4232	< 0,0001	

Tabela 31 – Comparação de médias e intervalo de confiança – VARIÁVEL Q31 – "Costuma tomar o café da manhã"

Considerando um nível de significância de 95% o teste de Tukey-Kramer evidencia que, as médias de cada grupo diferencia de zero e que para todas as variáveis, não existe diferença entre as médias dos grupos tratamento e controle. Os gráficos a seguir ajudarão à compreensão e reafirmação a conclusão tomada.

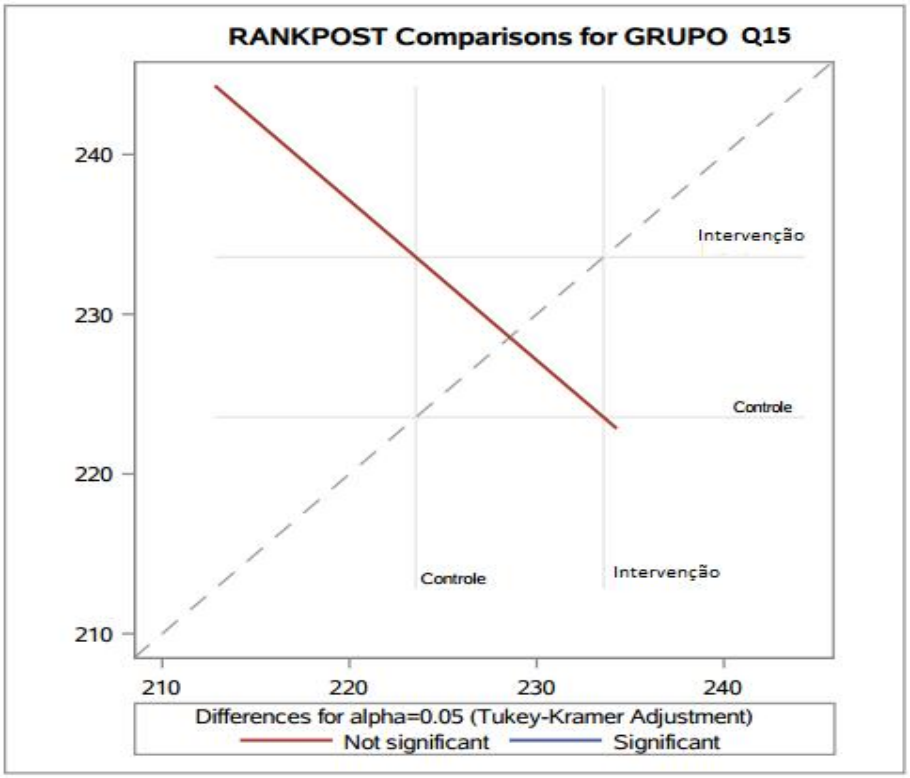

Figura 21 - Ajuste para múltiplas comparações das médias por Tukey – VARIÁVEL Q15

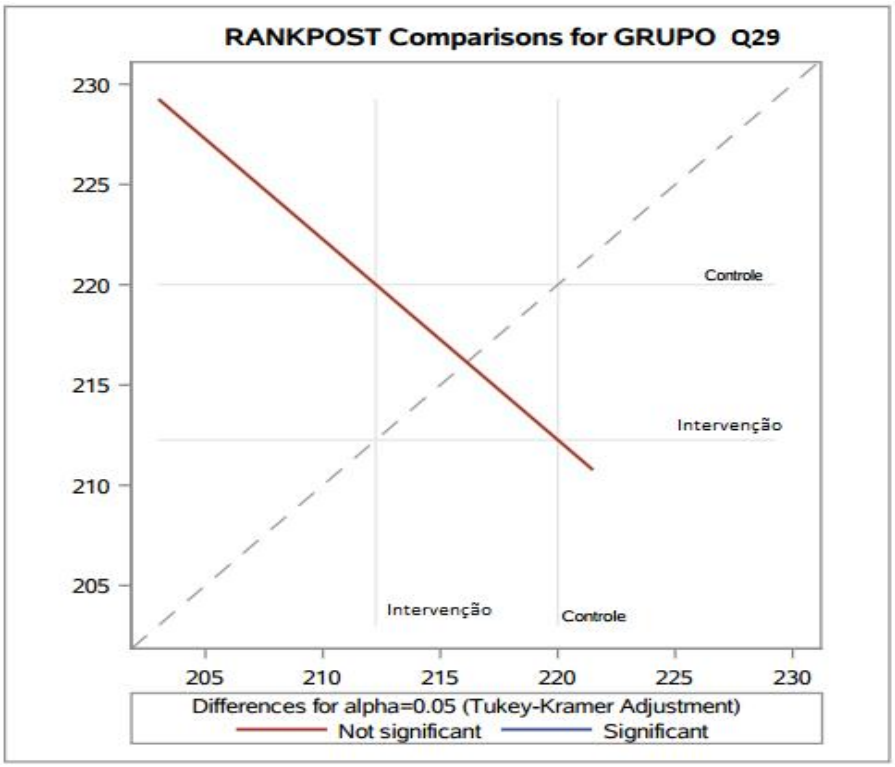

Figura 22 - Ajuste para múltiplas comparações das médias por Tukey – VARIÁVEL Q29

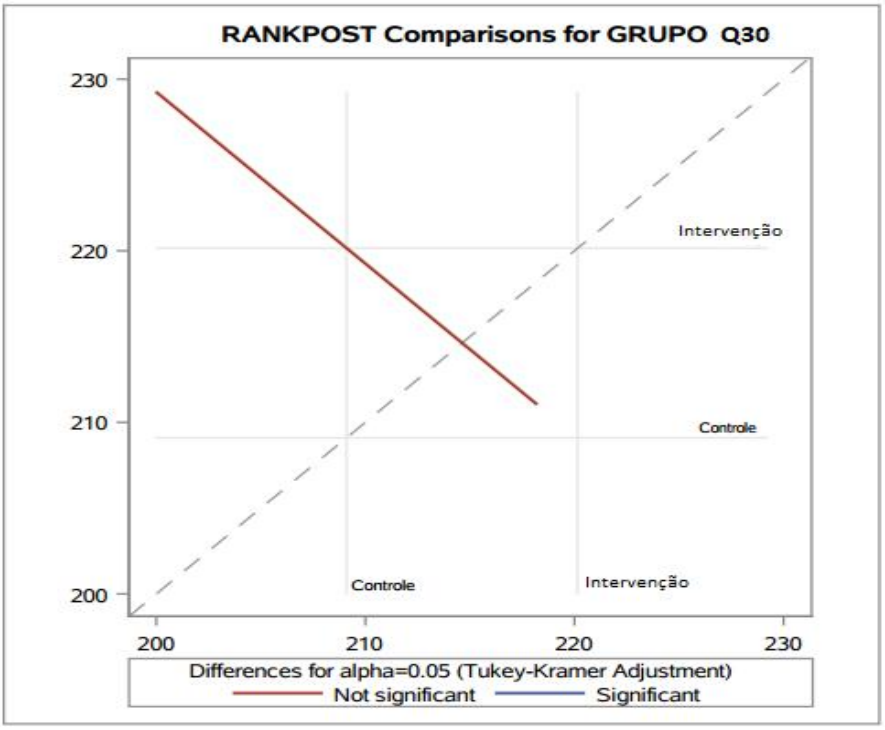

Figura 23 - Ajuste para múltiplas comparações das médias por Tukey – VARIÁVEL Q30

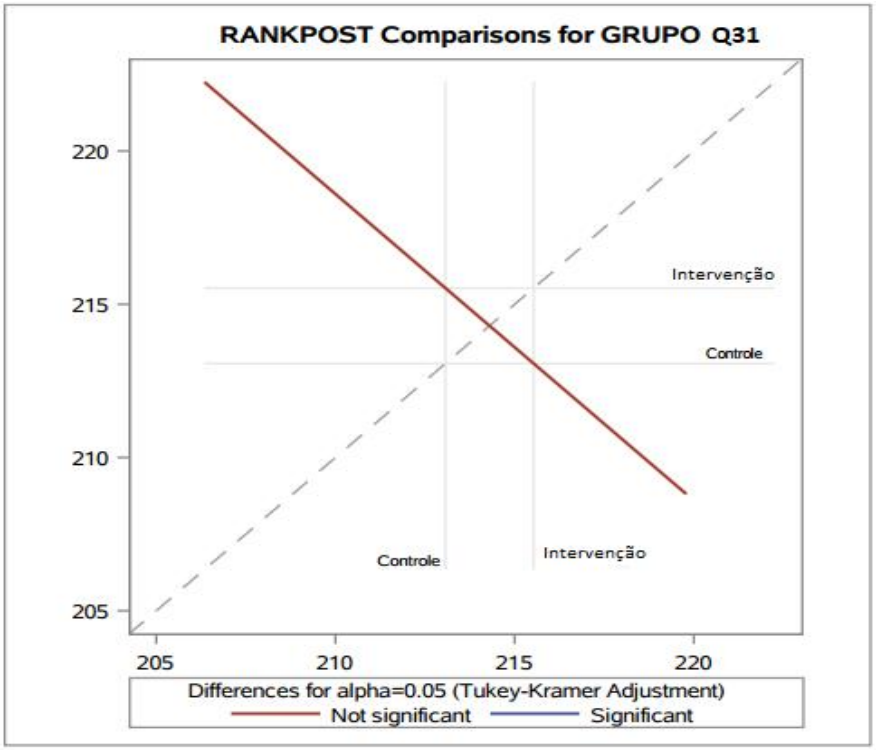

Figura 24 - Ajuste para múltiplas comparações das médias por Tukey – VARIÁVEL Q31

Assim como os demais, os testes para a diferença entre as médias do rankpre e rankpost para cada grupo não deram significativos. Como conclusão geral, para as quatro variáveis em questão, assumimos que há evidências que não existe diferença entre as médias do grupo controle e do grupo tratamento. Por fim, para diagnóstico do modelo, analisamos se a distribuição de probabilidade dos resíduos se aproxima da normal. Apesar de ser visível alguns pontos desviados da reta normal do "QQplot", assumimos uma convergência para uma distribuição Normal de probabilidade.

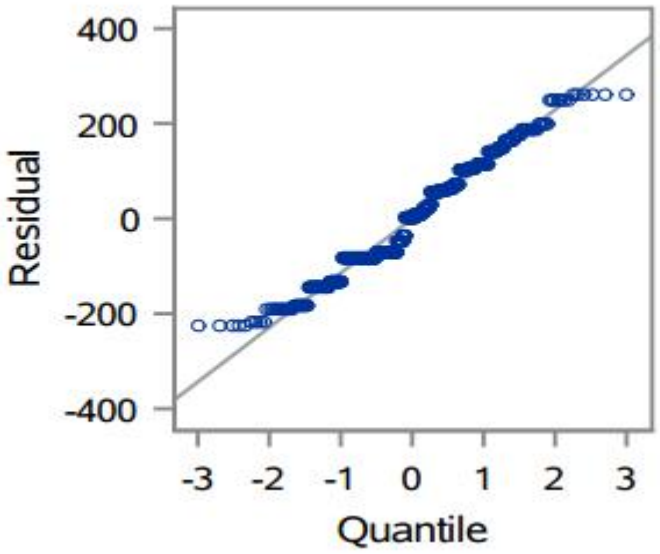

Figura 25 – Gráfico de probabilidade normal – VARIÁVEL 15

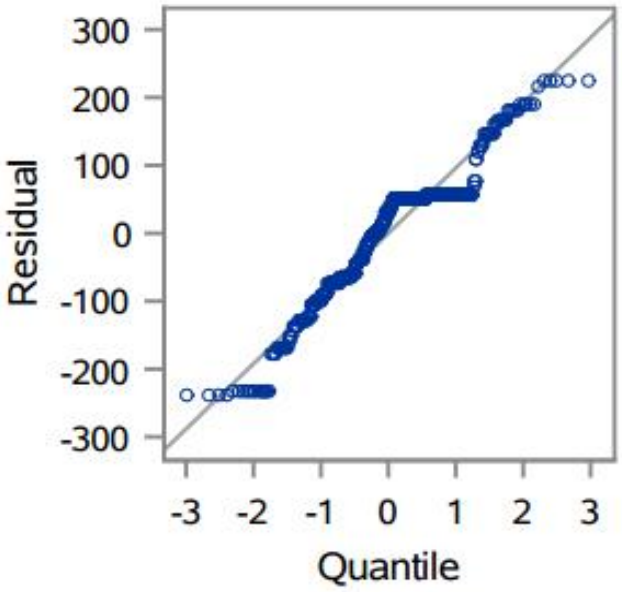

Figura 26 – Gráfico de probabilidade normal – VARIÁVEL 29

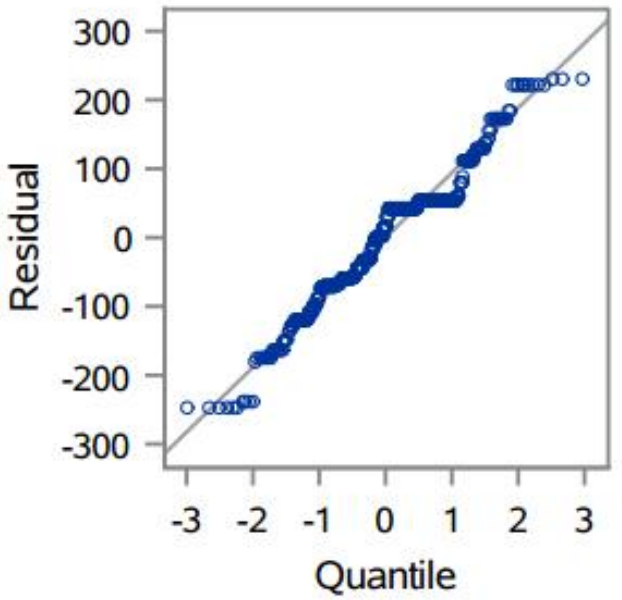

Figura 27 – Gráfico de probabilidade normal – VARIÁVEL 30

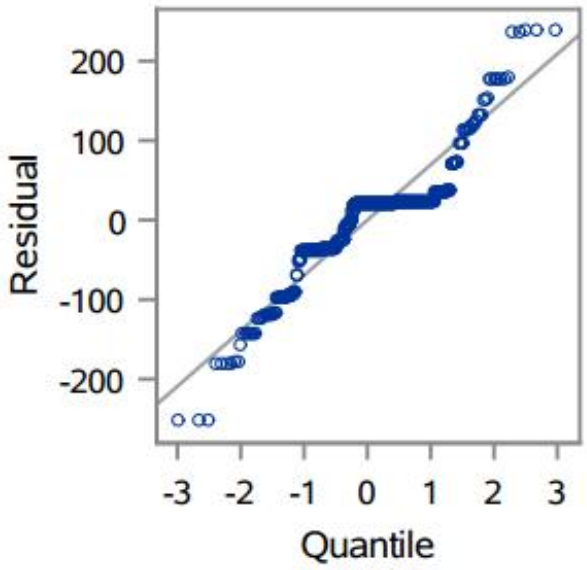

Figura 28 – Gráfico de probabilidade normal – VARIÁVEL 31

7 Considerações finais

Esse estudo teve por objetivo a aplicação de modelos de análise de covariância paramétrico e não paramétrico e a validade dos mesmos numa aplicação específica sobre conhecimento e vivência de hábitos alimentares saudáveis, cujos respondentes são alunos com idade entre 13 e 18 anos. Outro ponto importante é a verificação dos pressupostos.

Com os resultados apresentados no Capítulo 6, observou-se que para estudo de pré-teste e pós-teste os modelos ANCOVA são de grande valia e eficiência, uma vez que os resultados obtidos, bem como as suas interpretações, sobre o ajuste e aplicação do modelo apresentaram resultados mais consistentes uma vez que foi controlado pela fonte de variação do pré-teste.

A análise de covariância foi aplicada para verificar a relação entre a variável resposta pós-teste e a covariável, o pré-teste, com o objetivo de reduzir a variabilidade do termo erro e fazer um estudo mais poderoso para comparar os resultados efeitos dos tratamentos.

O teste de hipótese para o resultado do pós-teste foi rejeitado para todas as variáveis não paramétricas estudadas, indicando que não houve diferença significativa entre a médias dos tratamentos controle e intervenção. Já os testes paramétricos, para as questões sobre conhecimento, foram significativas, indicando um sucesso no tratamento de intervenção com os alunos.

Referências bibliográficas

1. AGRANONIK M.; MACHADO L. R. Analysis of covariance: Application to pulmonary function data, adjusted for age, Programa de Pós-Graduação - Saúde da Criança e do Adolescente, Universidade Federal do Rio Grande do Sul (UFRGS), 2011.

2. BONATE, L. B. Analysis of Pretest-Posttest Designs. Chapmam e Hall/CRC, 2000, capítulo 5.

3. CONOVER, W. J. Practical nonparametric statistics, 1980, p.373.

4. Example 39.4 Analysis of Covariance. Suport.SAS. Disponível em: http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_glm_sect049.htm. Acesso em: 25 de mai. de 2018.

5. KUTNER M. H. et al. Applied Linear Statistical Models, 2005, capítulo 22.

6. KIRK R.E., Experimental Design: Procedures for the Behavioral Sciences, Brooks/Cole, Belmont, CA, 1982.

7. MONTGOMERY D.C. Designs and Analysis of Experiments 12^{th} edition, New York, 2001, J. Wiley.

8. QUADE D., Rank analysis of covariance, J. Am. Stat. Assoc., 62, 1187, 1967.

Anexo

Questionário

Universidade de Brasília/Faculdade de Ciências da Saúde
Programa de Pós-Graduação em Nutrição Humana
Avaliação do impacto do uso de um instrumento imagético para a promoção de práticas alimentares saudáveis entre adolescentes de Brasília-DF

IDENTIFICAÇÃO DO ALUNO

1) Qual é o seu nome? _____
2) Qual é a sua escola? _____
3) Qual é a sua Série/ Turma? _____
4) Qual período você estuda? 1. ( ) Manhã 2. ( ) Tarde
5) Qual é o seu sexo? 1. ( ) Masculino 2. ( ) Feminino
6) Qual é a sua idade? _____anos Data do seu aniversário: ____/____/_____
7) Qual é a sua cor/raça? 1. ( ) Branca 2. ( ) Preta 3. ( ) Amarela 4. ( ) Parda 5. ( ) Indígena
8) Você mora com sua mãe? 1. ( ) Sim 1. ( ) Não
9) Você mora com seu pai? 1. ( ) Sim 2. ( ) Não
10) Contando com você, quantas pessoas moram na sua casa ou apartamento? _____

11) Qual o nível (grau) de ensino que sua mãe estudou ou estuda?

1. ( ) Minha mãe não estudou.
2. ( ) Minha mãe começou o ensino fundamental (ou 1º grau), mas não terminou.
3. ( ) Minha mãe terminou o ensino fundamental (ou 1º grau).
4. ( ) Minha mãe começou o ensino médio (ou 2º grau), mas não terminou.
5. ( ) Minha mãe terminou o ensino médio (ou 2° grau).
6. ( ) Minha mãe começou a faculdade (ensino superior), mas não terminou.
7. ( ) Minha mãe terminou a faculdade (ensino superior, inclusive pós graduação, mestrado e doutorado em curso ou terminado).
8. ( ) Não sei.

12) Qual o nível (grau) de ensino que seu pai estudou ou estuda?

1. ( ) Meu pai não estudou.

2. ( ) Meu pai começou o ensino fundamental (ou 1º grau), mas não terminou.
3. ( ) Meu pai terminou o ensino fundamental (ou 1º grau).
4. ( ) Meu pai começou o ensino médio (ou 2º grau), mas não terminou.
5. ( ) Meu pai terminou o ensino médio (ou 2° grau).
6. ( ) Meu pai começou a faculdade (ensino superior), mas não terminou.
7. ( ) Meu pai terminou a faculdade (ensino superior, inclusive pós graduação, mestrado e doutorado em curso ou terminado).
8. ( ) Não sei.

As próximas perguntas referem-se a sua alimentação. Leve em conta tudo o que você comeu em casa, na escola, na rua, em lanchonetes, em restaurantes ou em qualquer outro lugar. (Marque um X na sua resposta)

13) Ontem, em quais refeições você comeu salada crua? Exemplos: alface, tomate, cenoura, pepino, cebola, etc.

1. ( ) Não comi salada crua ontem
2. ( ) No almoço de ontem
3. ( ) No jantar de ontem
4. ( ) No almoço e no jantar de ontem

14) Ontem, em quais refeições você comeu legumes ou verduras cozidas, sem contar batata e aipim (mandioca/macaxeira)?

1. ( ) Não comi legumes nem verduras cozidos ontem
2. ( ) No almoço de ontem
3. ( ) No jantar de ontem
4. ( ) No almoço e no jantar de ontem

15) Ontem, quantas vezes você comeu frutas frescas?

1. ( ) Não comi frutas frescas ontem
2. ( ) Uma vez ontem
3. ( ) Duas vezes ontem
4. ( ) Três vezes ou mais ontem

NOS ÚLTIMOS 7 DIAS, em quantos dias você comeu/bebeu os seguintes alimentos? (Marque um X na sua resposta)

	Alimento (s)	Frequência do consumo nos ÚLTIMOS 7 DIAS							
16)	Feijão	Não comi	1 dia	2 dias	3 dias	4 dias	5 dias	6 dias	Todos os dias
17)	Salgados frito. Exemplo: batata frita (sem contar a batata de pacote) ou salgados fritos como coxinha de galinha, quibe frito, pastel frito, acarajé, etc.	Não comi	1 dia	2 dias	3 dias	4 dias	5 dias	6 dias	Todos os dias
18)	Hambúrguer, salsicha, mortadela, salame, presunto, *nuggets* ou linguiça	Não comi	1 dia	2 dias	3 dias	4 dias	5 dias	6 dias	Todos os dias
19)	Pelo menos um tipo de legume ou verdura crus ou cozidos. Exemplo: couve, tomate, alface, abóbora, chuchu, brócolis, espinafre, etc. Não inclua batata e aipim (mandioca/macaxeira).	Não comi	1 dia	2 dias	3 dias	4 dias	5 dias	6 dias	Todos os dias
20)	Salada crua. Exemplo: alface, tomate, cenoura, pepino, cebola, etc.	Não comi	1 dia	2 dias	3 dias	4 dias	5 dias	6 dias	Todos os dias
21)	Legumes ou verduras cozidos na comida, inclusive sopa? Exemplo: couve, abóbora, chuchu, brócolis, espinafre, cenoura, etc. Não inclua batata e aipim (mandioca/macaxeira).	Não comi	1 dia	2 dias	3 dias	4 dias	5 dias	6 dias	Todos os dias
22)	Biscoitos salgados ou bolachas salgadas	Não comi	1 dia	2 dias	3 dias	4 dias	5 dias	6 dias	Todos os dias
23)	Biscoitos doces ou bolachas doces	Não comi	1 dia	2 dias	3 dias	4 dias	5 dias	6 dias	Todos os dias
24)	Salgadinho de pacote ou batata frita de pacote	Não comi	1 dia	2 dias	3 dias	4 dias	5 dias	6 dias	Todos os dias
25)	Guloseimas (doces, balas, chocolates, chicletes, bombons ou pirulitos)	Não comi	1 dia	2 dias	3 dias	4 dias	5 dias	6 dias	Todos os dias
26)	Frutas frescas ou salada de frutas	Não comi	1 dia	2 dias	3 dias	4 dias	5 dias	6 dias	Todos os dias
27)	Leite (Excluir “leite” de soja e considerar, por exemplo, leite com café ou chocolate, vitamina, mingau)	Não comi	1 dia	2 dias	3 dias	4 dias	5 dias	6 dias	Todos os dias
28)	Refrigerante	Não comi	1 dia	2 dias	3 dias	4 dias	5 dias	6 dias	Todos os dias

29) Você costuma almoçar ou jantar - com sua mãe, pai ou responsável?
    1. ( ) Não
    2. ( ) Sim, todos os dias
    3. ( ) Sim, 5 a 6 dias por semana
    4. ( ) Sim, 3 a 4 dias por semana
    5. ( ) Sim, 1 a 2 dias por semana
    6. ( ) Sim, mas apenas raramente

30) Você costuma comer quando está assistindo à TV ou estudando?
    1. ( ) Não
    2. ( ) Sim, todos os dias
    3. ( ) Sim, 5 a 6 dias por semana
    4. ( ) Sim, 3 a 4 dias por semana
    5. ( ) Sim, 1 a 2 dias por semana
    6. ( ) Sim, mas apenas raramente

31) Você costuma tomar o café da manhã?
    1. ( ) Não
    2. ( ) Sim, todos os dias
    3. ( ) Sim, 5 a 6 dias por semana
    4. ( ) Sim, 3 a 4 dias por semana
    5. ( ) Sim, 1 a 2 dias por semana
    6. ( ) Sim, mas apenas raramente

32) **Você considera sua alimentação saudável?** Pensando nisso, que nota você daria para sua alimentação, sendo 0 nada saudável e 10 muito saudável? _____

33) **Quanto você sabe sobre alimentação saudável?** Pensando nisso, que nota você daria para seu conhecimento, sendo 0 não sei nada e 10 sei muito sobre alimentação saudável? ____

O QUE VOCÊ ACHA sobre as seguintes afirmações? (Marque um X na sua resposta)

34)	Alimentação saudável significa seguir uma dieta.	Verdadeiro	Falso	Eu não sei
35)	Alimentação saudável significa comer diversos alimentos de forma moderada.	Verdadeiro	Falso	Eu não sei
36)	Em uma dieta equilibrada, é possível consumir doces.	Verdadeiro	Falso	Eu não sei
37)	Bolacha recheada contém pouca gordura e muito açúcar.	Verdadeiro	Falso	Eu não sei
38)	Bebidas como suco de caixinha contém pouco açúcar e muita fruta.	Verdadeiro	Falso	Eu não sei
39)	Os alimentos *in natura* devem ser a base da nossa alimentação.	Verdadeiro	Falso	Eu não sei
40)	Frutas e verduras podem ser opções de alimentos minimamente processados.	Verdadeiro	Falso	Eu não sei
41)	Os ingredientes e métodos usados na fabricação de alimentos processados tornam os alimentos menos saudáveis (Ex. conserva de legumes).	Verdadeiro	Falso	Eu não sei
42)	Alimentos ultraprocessados são mais saudáveis que os alimentos minimamente processados.	Verdadeiro	Falso	Eu não sei
43)	Devemos evitar fazer compras em feiras livres, pois as feiras têm poucas opções de alimentos saudáveis.	Verdadeiro	Falso	Eu não sei
44)	Cozinhar em casa é uma prática considerada saudável, pois existem várias opções de alimentos congelados e temperos prontos.	Verdadeiro	Falso	Eu não sei
45)	A falta de tempo, o ambiente e a companhia podem influenciar na qualidade da alimentação.	Verdadeiro	Falso	Eu não sei
46)	De forma geral, as informações, orientações e mensagens das propagandas de alimentos que passam na TV são confiáveis e, portanto, podemos acreditar nelas.	Verdadeiro	Falso	Eu não sei

GABARITO QUESTIONÁRIO DE CONHECIMENTO

47)	Alimentação saudável significa seguir uma dieta.	Verdadeiro	**Falso**	Eu não sei
48)	Alimentação saudável significa comer diversos alimentos de forma moderada.	**Verdadeiro**	Falso	Eu não sei
49)	Em uma dieta equilibrada, é possível consumir doces.	**Verdadeiro**	Falso	Eu não sei
50)	Bolacha recheada contém pouca gordura e muito açúcar.	Verdadeiro	**Falso**	Eu não sei
51)	Bebidas como suco de caixinha contém pouco açúcar e muita fruta.	Verdadeiro	**Falso**	Eu não sei
52)	Os alimentos *in natura* devem ser a base da nossa alimentação.	**Verdadeiro**	Falso	Eu não sei
53)	Frutas e verduras podem ser opções de alimentos minimamente processados.	**Verdadeiro**	Falso	Eu não sei
54)	Os ingredientes e métodos usados na fabricação de alimentos processados tornam os alimentos menos saudáveis (Ex. conserva de legumes).	**Verdadeiro**	Falso	Eu não sei
55)	Alimentos ultraprocessados são mais saudáveis que os alimentos minimamente processados.	Verdadeiro	**Falso**	Eu não sei
56)	Devemos evitar fazer compras em feiras livres, pois as feiras têm poucas opções de alimentos saudáveis.	Verdadeiro	**Falso**	Eu não sei
57)	Cozinhar em casa é uma prática considerada saudável, pois existem várias opções de alimentos congelados e temperos prontos.	Verdadeiro	**Falso**	Eu não sei
58)	A falta de tempo, o ambiente e a companhia podem influenciar na qualidade da alimentação.	**Verdadeiro**	Falso	Eu não sei
59)	De forma geral, as informações, orientações e mensagens das propagandas de alimentos que passam na TV são confiáveis e, portanto, podemos acreditar nelas.	Verdadeiro	**Falso**	Eu não sei

Programação

Abaixo segue toda a programação SAS utilizada para trabalhar o banco de dados e gerar os resultados.

Importação do dados – SAS ondemand

```
FILENAME REFFILE '/home/luizpaulo_lp0/DADOS/dadostcc.xlsx';
PROC IMPORT DATAFILE=REFFILE
        DBMS=XLSX
        OUT=WORK.IMPORT;
        GETNAMES=YES;
RUN;
PROC CONTENTS DATA=WORK.IMPORT; RUN;
run;
```

Formatação

```
data a;set WORK.IMPORT;
pint = sum (p1int,p2int,p3int);
if pint = 3 then pint1=1;
else pint1 = 0;
if q15 = . then q151 = .;
else if q15 = 1 then q151 = 0;
else if q15 = 2 then q151 = 1;
else if q15 = 3 then q151 = 2;
else if q15 = 4 then q151 = 3;

if q15dois = . then q152 = .;
else if q15dois = 1 then q152 = 0;
else if q15dois = 2 then q152 = 1;
else if q15dois = 3 then q152 = 2;
else if q15dois = 4 then q152 = 3;

if q29 = . then q291 = .;
else if q29 = 1 then q291 = 0;
else if q29 = 6 then q291 = 1;
else if q29 = 5 then q291 = 2;
else if q29 = 4 then q291 = 3;
else if q29 = 3 then q291 = 4;
else if q29 = 2 then q291 = 5;

if q29dois = . then q292 = .;
else if q29dois = 1 then q292 = 0;
else if q29dois = 6 then q292 = 1;
else if q29dois = 5 then q292 = 2;
else if q29dois = 4 then q292 = 3;
```

```
else if q29dois = 3 then q292 = 4;
else if q29dois = 2 then q292 = 5;

if q30 = . then q301 = .;
else if q30 = 1 then q301 = 0;
else if q30 = 6 then q301 = 1;
else if q30 = 5 then q301 = 2;
else if q30 = 4 then q301 = 3;
else if q30 = 3 then q301 = 4;
else if q30 = 2 then q301 = 5;

if q30dois = . then q302 = .;
else if q30dois = 1 then q302 = 0;
else if q30dois = 6 then q302 = 1;
else if q30dois = 5 then q302 = 2;
else if q30dois = 4 then q302 = 3;
else if q30dois = 3 then q302 = 4;
else if q30dois = 2 then q302 = 5;

if q31 = . then q311 = .;
else if q31 = 1 then q311 = 0;
else if q31 = 6 then q311 = 1;
else if q31 = 5 then q311 = 2;
else if q31 = 4 then q311 = 3;
else if q31 = 3 then q311 = 4;
else if q31 = 2 then q311 = 5;

if q31dois = . then q312 = .;
else if q31dois = 1 then q312 = 0;
else if q31dois = 6 then q312 = 1;
else if q31dois = 5 then q312 = 2;
else if q31dois = 4 then q312 = 3;
else if q31dois = 3 then q312 = 4;
else if q31dois = 2 then q312 = 5;
run;

proc format;
value grupo 1 = 'Intervenção'
            2 = 'Controle';
run;
```

Análise descritiva

```
proc freq;
tables Grupo   P1int   P2Int   P3int   Q2 Q3 Q4 Q5 Q6 Q7 Q8 Q9 Q10 Q11 Q12;
run;

proc freq;
tables q13-q46 Nfin1;
run;

proc freq;
tables  Q13dois       Q14dois       Q15dois       Q16dois       Q17dois
        Q18dois       Q19dois       Q20dois       Q21dois       Q22dois
        Q23dois       Q24dois       Q25dois       Q26dois       Q27dois
        Q28dois       Q29dois       Q30dois       Q31dois       Q32dois
        Q33dois       Q34dois       Q35dois       Q36dois       Q37dois
        Q38dois       Q39dois       Q40dois       Q41dois       Q42dois
        Q43dois       Q44dois       Q45dois       Q46dois
Nfin2;
run;
```

Análise paramétrica – modelo ANCOVA

```
/* Variáveis Nfin1 Nfin2 */

proc means mean clm maxdec=2 data=a;
class grupo;
var Nfin1 Nfin2;
format grupo grupo.;
run;

/*****************************************************************/
proc glm data=a PLOTS=(DIAGNOSTICS RESIDUALS meanplot(CL CONNECT));
      class grupo ;
      model Nfin2=pint1 grupo Nfin1/solution;
      lsmeans grupo / stderr pdiff=all cl /*adjust=bon*/;
      /*means grupo;*/
                  output out=b p=estimate r=residuo;
                  format grupo grupo.;
                  run;
data b;set a;
if grupo = 1;
run;
proc ttest;
paired Nfin2*Nfin1;
run;
/*****************************************************************/
```

```
data c;set a;
if grupo = 2;
run;
proc ttest;
paired Nfin2*Nfin1;
run;
```

/* Variáveis Q32 Q32dois */

```
proc means mean clm maxdec=2 data=a;
class grupo;
var Q32 Q32dois;
format grupo grupo.;
run;

proc glm data=a PLOTS=(DIAGNOSTICS RESIDUALS meanplot(CL CONNECT));
        class grupo ;
        model Q32dois=pint1 grupo Q32/solution;
        lsmeans grupo / stderr pdiff=all cl /*adjust=bon*/;
        /*means grupo;*/
                output out=b p=estimate r=residuo;
                format grupo grupo.;
                run;

data b;set a;
if grupo = 1;
run;
proc ttest;
paired Q32dois*Q32;
run;

data c;set a;
if grupo = 2;
run;
proc ttest;
paired Q32dois*Q32;
run;
```

/* Variáveis Q33 Q33dois */

```
proc means mean clm maxdec=2 data=a;
class grupo;
var Q33 Q33dois;
format grupo grupo.;
run;

proc glm data=a PLOTS=(DIAGNOSTICS RESIDUALS meanplot(CL CONNECT));
```

```
class grupo ;
model Q33dois=pint1 grupo Q33/solution;
lsmeans grupo / stderr pdiff=all cl /*adjust=bon*/;
/*means grupo;*/
        output out=b p=estimate r=residuo;
        format grupo grupo.;
        run;

data b;set a;
if grupo = 1;
run;
proc ttest;
paired Q33dois*Q33;
run;

data c;set a;
if grupo = 2;
run;
proc ttest;
paired Q33dois*Q33;
run;
```

Análise não paramétrica – modelo ANCOVA

```
/* Variáveis Q15 Q15dois */

data b;set a;
if q152 = . or q151 = . then delete;
run;

proc iml;
use b;
read all var {grupo q152 q151};
rankpre = ranktie(q151);
rankpost = ranktie(q152);
devrankx = (rankpre-sum(rankpre)/nrow(rankpre));
devranky = (rankpost-sum(rankpost)/nrow(rankpost));
b = inv(devrankx` * devrankx) * devrankx` * devranky;
yhat = devrankx * b;
residual = devranky - yhat;
create resdata var {grupo q151 q152 rankpre
rankpost yhat residual};
append var {grupo q151 q152 rankpre
rankpost yhat residual};
run;
proc print data=resdata;
run;
```

```
proc means data=resdata mean clm maxdec=2;
class grupo;
var rankpre;
format grupo grupo.;
run;

title 'Quade test for Nonparametric Analysis of Covariance';
proc glm data=resdata;
class grupo;
model residual=grupo;
run;

title 'Parametric Analysis of Covariance on Ranks';
proc glm data=resdata PLOTS=(DIAGNOSTICS RESIDUALS meanplot(cl));
class grupo;
model rankpost = rankpre grupo;
lsmeans grupo/stderr pdiff=all cl out=b2;
format grupo grupo.;
run;

proc means mean clm maxdec=2 data=resdata;
class grupo;
var rankpre rankpost;
format grupo grupo.;
run;

data b;set resdata;
if grupo = 1;
dif = rankpost-rankpre;
run;
proc univariate;
var dif;
run;

data c;set resdata;
if grupo = 2;
dif = rankpost-rankpre;
run;
proc univariate;
var dif;
run;
```

/* Variáveis Q29 Q29dois */

```
data b;set a;
if q292 = . or q291 = . then delete;
run;

proc iml;
use b;
read all var {grupo q292 q291};
rankpre = ranktie(q291);
rankpost = ranktie(q292);
devrankx = (rankpre-sum(rankpre)/nrow(rankpre));
devranky = (rankpost-sum(rankpost)/nrow(rankpost));
b = inv(devrankx` * devrankx) * devrankx` * devranky;
yhat = devrankx * b;
residual = devranky - yhat;
create resdata var {grupo q291 q292 rankpre
rankpost yhat residual};
append var {grupo q291 q292 rankpre
rankpost yhat residual};
run;

proc print data=resdata;
run;

proc means data=resdata mean clm maxdec=2;
class grupo;
var rankpre;
format grupo grupo.;
run;

title 'Quade test for Nonparametric Analysis of Covariance';
proc glm data=resdata;
class grupo;
model residual=grupo;
run;

title 'Parametric Analysis of Covariance on Ranks';
proc glm data=resdata PLOTS=(DIAGNOSTICS RESIDUALS meanplot(cl));
class grupo;
model rankpost = rankpre grupo;
lsmeans grupo/stderr pdiff=all cl out=b2;
format grupo grupo.;
run;

proc means mean clm maxdec=2 data=resdata;
class grupo;
```

```
var rankpre rankpost;
format grupo grupo.;
run;

data b;set resdata;
if grupo = 1;
dif = rankpost-rankpre;
run;
proc univariate;
var dif;
run;

data c;set resdata;
if grupo = 2;
dif = rankpost-rankpre;
run;
proc univariate;
var dif;
run;
```

/* Variáveis Q30 Q30dois */

```
data b;set a;
if q302 = . or q301 = . then delete;
run;

proc iml;
use b;
read all var {grupo q302 q301};
rankpre = ranktie(q301);
rankpost = ranktie(q302);
devrankx = (rankpre-sum(rankpre)/nrow(rankpre));
devranky = (rankpost-sum(rankpost)/nrow(rankpost));
b = inv(devrankx` * devrankx) * devrankx` * devranky;
yhat = devrankx * b;
residual = devranky - yhat;
create resdata var {grupo q301 q302 rankpre
rankpost yhat residual};
append var {grupo q301 q302 rankpre
rankpost yhat residual};
run;

proc print data=resdata;
run;

proc means data=resdata mean clm maxdec=2;
class grupo;
```

```
var rankpre;
format grupo grupo.;
run;

title ’Quade test for Nonparametric Analysis of Covariance’;
proc glm data=resdata;
class grupo;
model residual=grupo;
run;

title ’Parametric Analysis of Covariance on Ranks’;
proc glm data=resdata PLOTS=(DIAGNOSTICS RESIDUALS meanplot(cl));
class grupo;
model rankpost = rankpre grupo;
lsmeans grupo/stderr pdiff=all cl out=b2;
format grupo grupo.;
run;

proc means mean clm maxdec=2 data=resdata;
class grupo;
var rankpre rankpost;
format grupo grupo.;
run;
data b;set resdata;
if grupo = 1;
dif = rankpost-rankpre;
run;
proc univariate;
var dif;
run;
data c;set resdata;
if grupo = 2;
dif = rankpost-rankpre;
run;
proc univariate;
var dif;
run;
```

/* Variáveis Q31 Q31dois */

```
data b;set a;
if q312 = . or q311 = . then delete;
run;

proc iml;
use b;
read all var {grupo q312 q311};
```

```
rankpre = ranktie(q311);
rankpost = ranktie(q312);
devrankx = (rankpre-sum(rankpre)/nrow(rankpre));
devranky = (rankpost-sum(rankpost)/nrow(rankpost));
b = inv(devrankx` * devrankx) * devrankx` * devranky;
yhat = devrankx * b;
residual = devranky - yhat;
create resdata var {grupo q311 q312 rankpre
rankpost yhat residual};
append var {grupo q311 q312 rankpre
rankpost yhat residual};
run;
proc print data=resdata;
run;

proc means data=resdata mean clm maxdec=2;
class grupo;
var rankpre;
format grupo grupo.;
run;

title 'Quade test for Nonparametric Analysis of Covariance';
proc glm data=resdata;
class grupo;
model residual=grupo;
run;

title 'Parametric Analysis of Covariance on Ranks';
proc glm data=resdata PLOTS=(DIAGNOSTICS RESIDUALS meanplot(cl));
class grupo;
model rankpost = rankpre grupo;
lsmeans grupo/stderr pdiff=all cl out=b2;
format grupo grupo.;
run;

proc means mean clm maxdec=2 data=resdata;
class grupo;
var rankpre rankpost;
format grupo grupo.;
run;
data b;set resdata;
if grupo = 1;
dif = rankpost-rankpre;
run;
proc univariate;
var dif;
run;
```

```
data c;set resdata;
if grupo = 2;
dif = rankpost-rankpre;
run;
proc univariate;
var dif;
run;
```