UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
DEPARTAMENTO DE ESTATÍSTICA

Análise de Experimentos com Repetição Única através do Coeficiente de Correlação Intraclasse

Autor: Gustavo Thomas
Orientadora: Professora Dra. Stela Castro

Porto Alegre, 15 de dezembro de 2015.

Universidade Federal do Rio Grande do Sul
Instituto de Matemática
Departamento de Estatística

Análise de Experimentos com Repetição Única através do Coeficiente de Correlação Intraclasse

Autor: Gustavo Thomas

Trabalho de Conclusão de Curso
apresentado para obtenção
do grau de Bacharel em Estatística.

Banca Examinadora:
Professora Dra. Stela Castro
Professor Dr. João Riboldi

Porto Alegre, 15 de dezembro de 2015.

Aos meus pais, Silvério e Claci, e aos meus irmãos, Rodrigo e Dionei.

“O aumento do conhecimento é como uma esfera dilatando-se no espaço: quanto maior a nossa compreensão, maior o nosso contato com o desconhecido.”
Blaise Pascal

Agradecimentos

Agradeço primeiramente aos meus pais, minha base, por aceitarem minhas escolhas desde que saí de casa. Obrigado por manterem contato quando estive longe, por comemorarem comigo nas vitórias e me animarem nos momentos difíceis.

Agradeço aos meus irmãos, vocês são uma espécie de pais e melhores amigos ao mesmo tempo. Obrigado Nei, sem você eu nem teria vindo pra cá. Obrigado Ro, pelos conselhos e apoio constante, e também cunhada Fabi, pela acolhida nos fins de semana e bons momentos que vivemos quando moraram aqui.

Agradeço a todos os professores com quem tive contato na graduação, principalmente aos professores do Departamento de Estatística, pelos ensinamentos e dedicação a cada um de nós. Um agradecimento especial à professora Stela, por me orientar neste trabalho e me aconselhar sobre o futuro.

Agradeço à UFRGS por proporcionar um ensino de qualidade e pela assistência estudantil ofertada, e ao governo pela oportunidade do intercâmbio de graduação através do programa Ciência sem Fronteiras.

Agradeço aos amigos que fiz aqui e que sei que posso contar quando e onde estiver, vocês fizeram esses anos passarem muito rápido. Também agradeço a todos que de uma ou de outra forma me auxiliaram nesse período.

Por fim, agradeço à Deus por me dar força, persistência e saúde para seguir em frente e tranquilidade e confiança para tomar as decisões que julguei corretas.

Resumo

Em muitos experimentos não é possível realizar repetições dos tratamentos nas unidades experimentais devido a questões financeiras e logísticas, entre outras. A falta de repetição impede que os métodos tradicionais de análise sejam utilizados. Porém, frequentemente há a disponibilidade de subunidades amostrais em tais experimentos, que permitem que um método alternativo de análise seja utilizado.

O objetivo desse trabalho é apresentar tal alternativa de análise, proposta por Perrett (2004), e demonstrar sua utilização em um experimento odontológico. O método faz uso de estimativas para o coeficiente de correlação intraclasse para aproximar o erro experimental, que não pode ser estimado da forma usual.

Através da comparação dos resultados obtidos com diferentes estimativas para a correlação intraclasse, foi possível observar que os resultados com esse método são bastante dependentes do valor da estimativa utilizada e que os testes tem seu poder reduzido conforme aumenta o valor da estimativa. Apesar disso, a aplicação do método é simples, e de posse de uma estimativa precisa para a correlação intraclasse ele possibilita a realização de análises válidas em experimentos com repetição única.

Sumário

1. Introdução

Na etapa do planejamento de um experimento, com frequência o experimentador ou pesquisador encontra restrições para realizar repetições dos tratamentos nas unidades experimentais. Experimentos com repetição única ocorrem em diversas áreas e por diferentes motivos. Podem-se citar os experimentos onde as unidades experimentais são destruídas, muito caras, de disponibilidade rara ou de difícil acesso. Ainda, podem ocorrer situações onde os testes são extremamente demorados ou muito custosos computacionalmente.

Lidar com a análise estatística desse tipo de situação é um tema que tem sido discutido na literatura desde os anos 80, segundo Hamada e Balakrishnan (1998). A falta de repetição implica na falta de variabilidade na resposta obtida em cada tratamento. Isso se traduz na impossibilidade de utilização dos métodos tradicionais de análise, pois não havendo mais de uma unidade experimental por tratamento não é possível estimar o erro experimental, que é necessário para o cálculo dos testes t de Student ou F de Snedecor (no caso múltiplo) comumente usados no caso de comparação de médias de tratamentos.

Em muitos experimentos não repetidos, porém, se tem a disponibilidade de subunidades amostrais para cada unidade não repetida. Por exemplo, considere a situação onde um pesquisador da área da educação quer comparar 2 métodos de ensino e tem acesso a apenas 2 salas de aula de uma escola. Dessa forma, cada método de ensino será aplicado a apenas uma sala de aula (repetição única), mas haverá várias medidas da variável resposta de acordo com o número de alunos (subunidades experimentais) por sala de aula (unidade experimental). Em outro cenário, um agrônomo pretende comparar o efeito de 2 métodos de plantio em terrenos com diferentes sistemas de irrigação. Por questões logísticas, não é possível replicar os sistemas de irrigação e cada terreno é subdividido com os métodos de plantio aleatoriamente alocados nas diferentes porções.

Nessas situações, a presença de subunidades experimentais permite uma forma de análise alternativa para experimentos sem repetição. Com elas, é possível calcular a variabilidade da variável resposta devida às subunidades, e, com uma estimativa do coeficiente de correlação intraclasse (CCI), pode-se substituir a variabilidade devida às unidades experimentais e realizar os testes tradicionais. Para tanto, há a necessidade de um número razoável de subunidades e de um bom conhecimento prévio da variabilidade da medida em estudo. É importante lembrar que essa abordagem é apenas uma forma de se realizar a análise em situações onde não é possível ou viável a repetição dos tratamentos, não descartando ou substituindo de forma alguma a necessidade de repetição quando esta for possível.

Ideias similares ao uso da correlação intraclasse vinham sendo estudadas desde a década de 80, mas a formalização do método se deu somente na década passada. Barcikowski (1981) debateu a importância de se considerar o grupo de subunidades como unidade experimental apesar da redução no poder estatístico e apresentou fórmulas que demonstram o tamanho amostral necessário para um desejado poder estatístico utilizando essa abordagem. Blair et al. (1983) revelaram como a escolha incorreta da unidade de análise aumenta a chance de se concluir que existem efeitos significativos de tratamento quando estes são nulos (erro Tipo I). O uso de uma estimativa *a priori* para o coeficiente de correlação intraclasse na análise de experimentos não replicados também foi abordada por Blair e Higgins (1986), que reportaram a ocorrência de um aumento no poder do teste, especialmente no caso de amostras pequenas. Mais recentemente, Bond e Higgins (2001) apresentaram o uso de estimadores bayesianos na obtenção de valores *a priori* para o CCI. Finalmente, em 2004, J. J. Perrett caracterizou e apresentou formalmente o método do uso de um valor aproximado do CCI para a análise de experimentos com repetição única, discutindo ainda a utilização de distribuições e a experiência do pesquisador na aproximação do valor do CCI.

Para a obtenção de estimativas do CCI, Perrett propõe 4 estratégias. As duas primeiras se baseiam na experiência do pesquisador, que define um valor máximo ou um intervalo (valores máximo e mínimo) que considera razoável para o CCI no experimento, a partir dos quais se realizam os testes e se conclui para cada caso. Caso o pesquisador não tenha conhecimento prévio sobre o CCI, mas tenha uma ideia da sua distribuição, uma alternativa é definir uma distribuição *a priori* para o mesmo e usar o valor esperado dessa para os testes. Como uma segunda alternativa, Perrett argumenta também que se pode buscar estimativas para o coeficiente em estudos similares na literatura, utilizando a mediana dos valores encontrados ou realizar os testes com cada um deles e depois fazer uma média ponderada dos p-valores encontrados.

Apesar de ser bastante versátil e simples, a técnica ainda é pouco conhecida e utilizada. Poucos experimentos publicados foram analisados através da técnica até então, dentre os quais se pode citar Perrett e Higgins (2006) e Machado et al. (2013). Dessa forma, o objetivo desse trabalho é apresentar esse método de análise de experimentos de repetição única e demonstrar o seu uso através da análise de um experimento odontológico.

2. Método e aplicação

2.1 Modelo experimental

Para introduzir o contexto do experimento que será analisado neste trabalho, considere a seguinte situação. Um pesquisador deseja comparar 4 lubrificantes de parafusos em implantes dentários através da força de fixação do parafuso ao implante (pré-carga), medida em Kg, com o objetivo de encontrar um lubrificante que resulte na maior fixação média entre os 4 testados. Como o único fator controlado no experimento é o lubrificante, trata-se de um delineamento de tratamentos unifatorial. Quanto à aleatorização, não há restrições para alocar cada implante a um determinado lubrificante, o que caracteriza o delineamento do experimento como completamente casualizado.

Seja i=1,2,...,t o índice do lubrificante e j=1,2,...,r o índice do conjunto implante/parafuso. Caso o experimento tivesse *r* repetições de conjuntos implante/parafuso por lubrificante, ter-se-ia o seguinte modelo estatístico para o experimento:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \tag{2.1.1}$$

onde:

Y_{ij} representaria a pré-carga observada (em Kg) no conjunto implante/parafuso *j* submetido ao lubrificante *i*;

μ representaria a média geral, comum a todos os lubrificantes;

τ_i representaria o efeito fixo do lubrificante *i*;

ε_{ij} representaria o erro aleatório associado à pré-carga observada no conjunto implante/parafuso *j* submetido ao lubrificante *i*.

Essa seria a situação ideal, onde o erro experimental, que é variação intrínseca das unidades experimentais submetidas ao mesmo tratamento, poderia ser estimado com t(r-1) graus de liberdade. A estatística do teste (ET) para o modelo (2.1.1), no caso de comparação de duas médias de tratamentos, seria dada pela fórmula (2.1.2) abaixo, onde $\bar{Y}_{1.}$ e $\bar{Y}_{2.}$ denotam as duas médias de tratamentos e r_1 e r_2 denotam o número de observações em cada tratamento.

$$ET = \frac{\bar{Y}_{1.} - \bar{Y}_{2.}}{\sqrt{\sigma^2{}_{\varepsilon} * \left\{\frac{1}{r_1} + \frac{1}{r_2}\right\}}} \tag{2.1.2}$$

Entretanto, a situação real é que o pesquisador tem acesso a apenas 4 implantes, onde deverá fixar todos os parafusos disponíveis. O que é feito nesse caso é dividir o número de parafusos disponíveis em 4 grupos, um grupo de parafusos para cada implante. Assim, cada lubrificante será aplicado a vários parafusos, mas que serão testados em apenas um implante por lubrificante. Dessa forma, haverá várias medidas por tratamento de acordo com o número de parafusos testados, mas apenas um implante por tratamento, ou seja, os parafusos podem ser considerados como subunidades de cada implante, que é a verdadeira unidade experimental (e que não é repetida). Sendo k=1,2,...,p o índice adotado para designar o parafuso, o modelo estatístico para o experimento real é dado por:

$$Y_{ik} = \mu + \tau_i + \beta_i + \xi_{ik} \tag{2.1.3}$$

onde:

Y_{ik} representa a pré-carga observada (em Kg) ao fixar o parafuso *k* no implante submetido à forma de lubrificação *i*;

μ representa a média geral, comum a todos os lubrificantes;

τ_i representa o efeito fixo do lubrificante *i*;

β_i representa o efeito aleatório do implante onde serão fixados somente parafusos que receberam o lubrificante *i*;

ξ_{ik} representa o efeito aleatório do parafuso *k* no implante sob a lubrificante *i*.

Como só há um implante por tratamento foi suprimido o índice de implante, que seria idêntico ao índice *i* do lubrificante. A principal diferença entre os dois modelos é a divisão do erro aleatório tradicional (ε_{ij}) em (2.1.1) em dois termos distintos em (2.1.3), sendo um para a variabilidade entre diferentes unidades experimentais (implantes) e outro para a variabilidade entre as subunidades experimentais (parafusos) em cada unidade experimental (isto é, dentro de unidades experimentais). Definidos os termos, tem-se que os pressupostos exigidos para a realização dos testes de hipóteses sobre os parâmetros de interesse podem ser enumerados como segue.

1. Os efeitos aleatórios de implante e parafuso são independentes, ou ainda, $Cov(\beta_i,\xi_{ik})=0$;
2. O efeito aleatório de implante é normalmente distribuído, com média nula e variância constante, ou ainda, $\beta_i \sim N(0,\sigma^2_\beta)$;
3. O efeito aleatório de parafuso é normalmente distribuído, com média nula e variância constante, ou ainda, $\xi_{ik} \sim N(0,\sigma^2_\xi)$.

De forma análoga à expressão (2.1.2), neste caso a estatística do teste para a comparação de duas médias de tratamentos é dada pela expressão (2.1.4) abaixo, onde e p_1

e p_2 denotam o número de observações (subunidades) em cada tratamento $p^2{}_1$ e $p^2{}_2$ denotam o quadrado do número de observações em cada tratamento.

$$ET = \frac{\bar{Y}_{1.} - \bar{Y}_{2.}}{\sqrt{\sigma^2{}_\xi * \left\{\frac{1}{p_1} + \frac{1}{p_2}\right\} + \sigma^2{}_\beta * \left[\left(\frac{1}{p_1}\right)^2 * p^2{}_1 + \left(\frac{1}{p_2}\right)^2 * p^2{}_2\right]}} \quad (2.1.4)$$

A expressão da estatística do teste para comparação de mais de duas médias de tratamentos pode ser encontrada em Perrett (2004), página 50. Comparando (2.1.4) com a expressão (2.1.2), pode-se observar que ao considerar os parafusos como unidades experimentais não é levada em conta (assumida como nula) a variabilidade entre implantes, representada pelo termo extra no denominador de (2.1.4). Dessa forma, o modelo (2.1.1) tem sua ET e consequentemente seu erro Tipo I inflacionados em relação ao modelo corretamente especificado (2.1.3) que considera os implantes como unidades experimentais.

Para proceder com a análise do modelo (2.1.3), é necessário conhecer a variância da pré-carga entre implantes ($\sigma^2{}_\beta$) e a variância da pré-carga entre parafusos ($\sigma^2{}_\xi$), que são as duas fontes de variabilidade no denominador de (2.1.4). Como se dispõe de várias medidas para cada tratamento (uma para cada parafuso), $\sigma^2{}_\xi$ pode ser estimado de forma bastante precisa por uma variância ponderada por implante. Entretanto, não é possível mensurar $\sigma^2{}_\beta$ da forma usual, visto que cada lubrificante foi testado em apenas um implante, ou seja, só se tem a medida de uma unidade experimental distinta por tratamento (repetição única). Como alternativa, Perrett (2004) propõe, pela relação entre $\sigma^2{}_\beta$ e $\sigma^2{}_\xi$ dada pelo CCI, uma forma prática de se estimar $\sigma^2{}_\beta$ de forma indireta, apresentada na seção 2.2.

2.2 Uso do Coeficiente de Correlação Intraclasse para estimar $\sigma^2{}_\beta$

O coeficiente de correlação intraclasse (usualmente denotado por ρ) foi introduzido por Sr. Ronald Fisher (1925) e é definido pela seguinte expressão:

$$\rho = \frac{Cov(Y_{ijk}, Y_{ijk'})}{\sqrt{Var(Y_{ijk}) + Var(Y_{ijk'})}} = \frac{\sigma^2{}_\beta}{\sqrt{(\sigma^2{}_\beta + \sigma^2{}_\xi)(\sigma^2{}_\beta + \sigma^2{}_\xi)}} = \frac{\sigma^2{}_\beta}{\sigma^2{}_\beta + \sigma^2{}_\xi} \quad (2.2.1)$$

Por (2.2.1) pode-se observar que o valor do coeficiente será próximo de zero caso a variabilidade entre unidades experimentais for muito menor que a variabilidade dentro de unidades experimentais (ou entre subunidades) e próximo da unidade caso contrário, sendo que zero e um são justamente os valores mínimo e máximo que o coeficiente assume. Entretanto, o termo de interesse em (2.2.1) é $\sigma^2{}_\beta$, que por sua vez pode ser expresso por:

$$\sigma^2{}_\beta = \sigma^2{}_\xi * \frac{\rho}{(1-\rho)} \qquad (2.2.2)$$

Assim, tem-se que a variância entre implantes pode ser estimada a partir da variância entre parafusos e de uma estimativa para o CCI. Aplicando (2.2.2) na ET (2.1.4), obtém-se:

$$ET = \frac{\bar{Y}_{1.} - \bar{Y}_{2.}}{\sqrt{\sigma^2{}_\xi * \left\{\left(\frac{\rho}{(1-\rho)}\right) * \left[\left(\frac{1}{p_1}\right)^2 * p^2{}_1 + \left(\frac{1}{p_2}\right)^2 * p^2{}_2\right] + \left[\left(\frac{1}{p_1}\right)^2 + \left(\frac{1}{p_2}\right)^2\right]\right\}}} \qquad (2.2.3)$$

De posse de uma estimativa para o CCI (ou de mais de uma), pode-se empregar a estratégia mais conveniente e segura dentre as citadas na introdução para testar a hipótese de pesquisa através de (2.2.3). Na sequência, são descritas mais detalhadamente as estratégias utilizadas nesse trabalho. Para maiores informações sobre todas as estratégias, ver Perrett (2004).

2.3 Estratégias para a obtenção de estimativas para ρ

Estimativas para o parâmetro ρ podem ser obtidas essencialmente de 2 formas: pontualmente, através do conhecimento do pesquisador sobre a variabilidade dos fatores experimentais ou através de estudos anteriores similares onde se consiga obter estimativas da variabilidade dos fatores; ou de forma variável, onde se assume uma distribuição de probabilidades para ρ que se considere mais adequada e se usa o valor esperado da mesma como estimativa para o coeficiente.

Para cada caso, Perrett (2004) faz algumas recomendações e observações, resumidas como segue.

Caso ρ fixo (estimado pontualmente):

- Deve-se ter variabilidade entre unidades experimentais (implantes) menor ou igual à variabilidade dentro das unidades (ou seja, $\rho \leq 0.5$). Caso contrário, o teste ficará com poder muito baixo para detectar diferenças;
- Para que o teste tenha nível de significância (α) real menor ou igual nível estabelecido, ρ escolhido deve ser maior ou igual ao ρ real;
- Em contrapartida, caso seja utilizado ρ pontual menor do que o real resultará em um teste com poder maior para detectar diferenças, mas ocasionará também uma inflação na probabilidade de erro do tipo I (aumento do α real);

- Há pouca diferença nos níveis de significância, e consequentemente no poder do teste, ao se realizar o experimento com 10 subunidades (parafusos) por unidade experimental ou um número infinito de subunidades;

Caso ρ variável (estimado através de uma distribuição de probabilidades):

- É utilizado o valor esperado da distribuição escolhida na realização dos testes;
- A distribuição assumida para ρ deve conter todo o conhecimento que se tenha sobre a variabilidade do coeficiente;
- O autor indica 2 formas de uso do número de graus de liberdade do erro para o caso de ρ variável:
    - Usando os graus de liberdade totais do erro, dados pelo número de tratamentos vezes o tamanho amostral decrescido de 1 unidade, que costuma resultar em níveis de significância reais um pouco acima dos nominais mas com bons níveis de poder;
    - Através do ajuste de Satterthwaite, que no caso de experimentos sem repetição fará com que o nível de significância não ultrapasse o estabelecido, mas reduzirá bastante o poder do teste.
- Por fim, Perrett (2004) defende que encontrar estimativas mais precisas para ρ é mais eficiente do que buscar um ajuste ótimo para os graus de liberdade do erro.

Para encontrar o número de graus de liberdade pelo ajuste de Satterthwaite, é preciso calcular uma variável auxiliar W, definida por

$$\mathrm{W} = \frac{\rho}{(1-\rho)} + \frac{1}{p} \tag{2.3.1}$$

e calcular a média e a variância desse novo conjunto de dados. De posse desses valores, Perrett (2004) demonstra que o número de graus de liberdade pelo ajuste de Satterthwaite para testes com t tratamentos é dado por

$$\mathrm{g.l.} = \frac{2\mathrm{t}(\mathrm{p}-1)[\mathrm{E(W)}]^2}{(\mathrm{tp}-\mathrm{t}+2) * \mathrm{Var(W)} + [\mathrm{E(W)}]^2} \tag{2.3.2}$$

onde E(W) representa o valor esperado e Var(W) a variância da variável auxiliar. O número de graus de liberdade obtido por (2.3.2) é então utilizado nos testes de comparação de médias de tratamentos.

2.4 Análise do experimento odontológico

Para ilustrar o método descrito por Perrett (2004), o experimento odontológico será analisado através de estimativas pontuais para ρ e assumindo uma distribuição de probabilidades para o mesmo. Adicionalmente, serão comparados ainda os resultados com $\hat{\rho}$ variável através da análise com uso dos graus de liberdade do erro totais e com o ajuste de Satterthwaite.

Na execução do experimento foram utilizados 28 parafusos divididos em 4 grupos, sendo 7 para cada implante/lubrificante. Para comparar as médias de pré-carga obtidas com os 4 lubrificantes, será realizada a técnica de análise de variância com nível de significância de 5%. Quando esta indicar diferenças significativas entre médias de tratamentos, a complementação será realizada pelo teste de Tukey, também a 5% de significância.

A estatística do teste de Tukey realizando a análise através do método de Perrett (2004) para comparação das médias de dois tratamentos 1 e 2, como o próprio descreve, é dada por:

$$ET_{Tukey} = \sqrt{2} * máx_{1,2} \frac{|\bar{Y}_1 - \bar{Y}_2|}{\sqrt{2\hat{\sigma}^2{}_{\xi} * \left\{\frac{\hat{\rho}}{1-\hat{\rho}} + \frac{1}{p}\right\}}} \qquad (2.4.1)$$

onde p denota o número de subunidades por unidade experimental. A diferença é dita significativa ao nível α de significância caso o valor de (2.4.1) exceda o quantil 1-α da distribuição da amplitude *studentizada* com o número de graus de liberdade do erro e número de tratamentos do experimento.

O código base para as análises está no Anexo I, que é basicamente o mesmo utilizado em todas as análises, sendo alterada somente a estimativa de ρ e o método de ajuste dos graus de liberdade no caso do ajuste de Satterthwaite. Para a realização das análises foi utilizado o software SAS University Edition com o auxílio do software R versão 3.2.0.

3. Resultados

3.1 Estimativas para ρ

Suponha que, por estudos odontológicos similares, tenha-se estimado que a correlação da pré-carga entre parafusos de um mesmo implante é de aproximadamente 0.3. Entretanto, suponha que o pesquisador que está conduzindo o experimento acredite, pela sua experiência na área, que essa correlação seja de apenas 0.1. Dessa forma, já se tem disponíveis duas estimativas pontuais para ρ: 0.1 e 0.3. Adicionalmente, será utilizada ainda a estimativa para ρ de 0.5, acima da qual Perrett (2004) considera impróprio o uso do método por resultar em um poder muito baixo para o teste; e a situação onde se ignora a correlação entre parafusos e os considera como unidades experimentais, isto é, usando $\hat{\rho}=0$. Caso o pesquisador não estivesse seguro em utilizar 0.1 como estimativa pontual para ρ, mas acreditasse que o valor real do parâmetro fosse menor que 0.3, poder-se-ia ainda atribuir uma distribuição de probabilidades a *priori* para o parâmetro, como a distribuição Beta com parâmetros α=5 e β=30, por exemplo, cuja função densidade de probabilidade é apresentada na Figura 3.1.1.

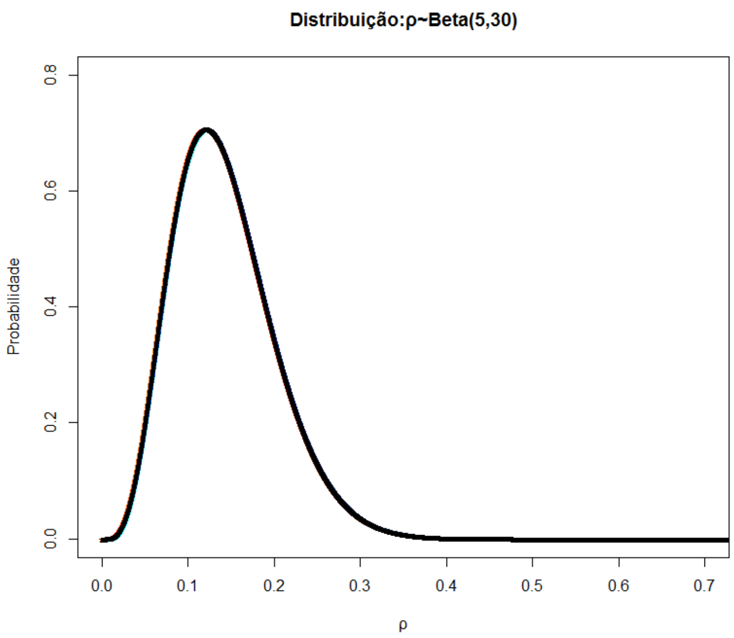

Figura 3.1.1 Distribuição a priori para ρ, Beta(5,30).

Pela Figura 3.1.1, pode-se observar que a distribuição de probabilidades escolhida atribui probabilidades maiores para valores de ρ entre 0.1 e 0.2, com probabilidades muito baixas de ρ estar acima de 0.3, contemplando o suposto conhecimento do pesquisador. O valor esperado dessa distribuição, que será uma das estimativas de ρ usada nas análises, é aproximadamente 0.1429 e o número de graus de liberdade pelo ajuste de Satterthwaite (por 2.3.2) vale aproximadamente 4.73.

3.2 Análises

Para a realização das análises de variância, foi verificado, através da análise de resíduos, que os pressupostos do modelo foram atendidos.

A seguir são apresentados os resultados da análise de variância e comparações múltiplas pelo teste de Tukey para cada estimativa de ρ, como descrito na seção 3.1.

Tabela 3.2.1 Análise de variância para cada valor de $\hat{\rho}$.

Causa da variação	$\hat{\rho}$	gl tratamento	gl do erro	Valor F	P-valor
Lubrificante	0[1]	3	24	138.09	< 0.0001
	0.1	3	24	40.09	< 0.0001
	0.3	3	24	13.73	< 0.0001
	0.5	3	24	6.30	0.0026
	0.1429	3	24	29.97	< 0.0001
	0.1429	3	4.73[2]	27.54	0.002

[1] Equivalente a considerar o parafuso como unidade experimental.
[2] Graus de liberdade obtidos pelo ajuste de Satterthwaite.

Tabela 3.2.2 Teste de Tukey para comparação de médias por lubrificante para cada valor de $\hat{\rho}$.

Lubrificante	Pré-carga média (kg)	$\hat{\rho} = 0$	$\hat{\rho} = 0.1$	$\hat{\rho} = 0.3$	$\hat{\rho} = 0.5$	$\hat{\rho} = 0.1429$[1]	$\hat{\rho} = 0.1429$[2]
Lubrificante 2	47.29	a	a	a	a	a	a
Lubrificante 4	42.20	b	a	ab	ab	ab	ab
Lubrificante 3	32.37	c	b	b	ab	bc	b
Lubrificante 1	31.15	c	b	b	b	c	b

Lubrificantes com letras iguais não diferem significativamente a 5% de significância.
[1] Análise com uso de graus de liberdade totais.
[2] Análise com uso do ajuste de Satterthwaite para os graus de liberdade.

Pelos resultados apresentados na Tabela 3.2.1, pode-se observar que para todas as estimativas de ρ houve rejeição da hipótese de igualdade das médias da pré-carga com o uso dos 4 lubrificantes a 5% de significância, apesar de haverem decréscimos consideráveis no valor do teste F ao aumentar o valor de $\hat{\rho}$. O valor F extremamente alto para o caso de $\hat{\rho}$=0 é resultado da inflação da estatística do teste ao se ignorar a variabilidade entre implantes, como mencionado na seção 2.1.

Os resultados dos testes de comparações múltiplas para as análises realizadas são apresentados na Tabela 3.2.2, onde se observa a mudança dos resultados dependendo do valor de $\hat{\rho}$ utilizado. Enquanto que há diferenças significativas entre as pré-cargas médias com qualquer combinação dos lubrificantes 3 e 1 com os lubrificantes 2 e 4 dois a dois ao usar $\hat{\rho} = 0.1$, com o uso de $\hat{\rho} = 0.5$ a única diferença significativa resultante é entre a pré-carga média observada com o lubrificante 2 e o lubrificante 1.

Com relação ao ajuste dos graus de liberdade por Satterthwaite, visualiza-se que ao fazer uso do mesmo há uma ligeira redução no valor do teste F em relação ao uso de graus de liberdade totais. Além disso, observa-se ainda que houve mudança no resultado da comparação do lubrificante 1 em relação ao lubrificante 4, onde a pré-carga média com este deixou de ser significativamente maior que aquele com o uso do ajuste por Satterthwaite.

3.3 Discussão

Pelas análises apresentadas, pode-se perceber que o valor da estimativa de $\hat{\rho}$ utilizada nas análises interfere diretamente nos resultados e consequentemente nas conclusões tiradas a partir deles. Caso o pesquisador usasse $\hat{\rho} = 0.1$ nas análises, chegaria à conclusão de que os lubrificantes 2 e 4 seriam igualmente mais indicados para aumentar a pré-carga, visto que pelas comparações múltiplas esses dois tratamentos apresentam pré-cargas médias significativamente maiores que os demais e não diferem entre si. Por outro lado, caso fosse utilizada a estimativa de correlação intraclasse de 0.3, 0.5 ou caso a correlação entre parafusos fosse ignorada, a recomendação seria fazer uso apenas do lubrificante 2. A provável causa para essa mudança nos resultados é o fato de que o aumento do valor de $\hat{\rho}$ acarreta uma redução no poder do teste, conforme descreve Perrett (2004). Além do mais, utilizando a estimativa de 0.5 para a correlação intraclasse observou-se que o número de diferenças significativas caiu de 4 para 1 (em comparação com $\hat{\rho} = 0.1$) pelo mesmo motivo, assim como um aumento no p-valor da análise de variância, ainda que sem alterar a conclusão.

A análise do experimento com o ajuste dos graus de liberdade por Satterthwaite também teve o efeito de diminuir o número de diferenças significativas encontradas nas comparações múltiplas em comparação à análise com uso de graus de liberdade totais. O ajuste de Satterthwaite tem o intuito de controlar o nível de significância do teste, de modo a assegurar que a probabilidade de erro tipo I não ultrapasse o nível α especificado. Por outro lado, tem como consequência a redução no poder do teste, sendo que Perrett (2004) classifica o uso do ajuste de Satterthwaite na análise como uma forma conservadora, para situações onde se quer mais segurança quanto às diferenças encontradas.

Como limitação dessa análise pode-se citar justamente a falta de conhecimento sobre o valor real do CCI. Por exemplo, a estimativa de 0.1 usada pode estar subestimando o valor real do coeficiente, sendo que essa análise poderia ser comparada à análise supondo independência entre as subunidades (considerando cada parafuso como uma unidade experimental), a qual se sabe não ser a abordagem correta. De fato, caso a correlação intraclasse verdadeira seja alta, é possível que mesmo as maiores diferenças observadas não sejam significativas; isto é, elas podem ser meramente devidas à mudança do implante e não à mudança do lubrificante. Por outro lado, superestimar o real valor do coeficiente de correlação intraclasse pode reduzir substancialmente o poder do teste, alterando o resultado das comparações múltiplas (como visto ao utilizar $\hat{\rho} = 0.5$) ou até mesmo conduzindo a um erro do tipo II.

4. Conclusão

Este trabalho teve como objetivo apresentar e demonstrar o uso do método de análise de experimentos com repetição única proposto por Perrett (2004) em um experimento com dados reais. A descrição do método foi feita no contexto do experimento odontológico analisado, destacando as principais diferenças dessa forma de análise em relação à abordagem incorreta das subunidades como unidades experimentais. Além disso, foram estabelecidas as duas formas de obtenção de estimativas para ρ e feitas as principais recomendações sobre cada uma segundo Perrett (2004).

Para a análise, foram utilizadas quatro estimativas pontuais de ρ e uma estimativa baseada no valor esperado da distribuição Beta (5,30), sendo que nesse último caso a análise com o uso de graus de liberdade totais foi comparada com a análise com o ajuste dos graus de liberdade por Satterthwaite. A partir dos resultados, foi possível verificar que as conclusões mudaram de acordo com a estimativa de ρ utilizada e com a redução dos graus de liberdade no sentido da diminuição do número de diferenças significativas em virtude da redução do poder do teste.

De uma forma geral, observou-se que a aplicação do método é bastante simples e permitiu realizar análises válidas para o experimento odontológico, onde não se tinha repetições genuínas e não era possível utilizar os métodos de análise tradicionais. Todavia, pôde-se perceber que o ponto fraco desse método é a sua forte dependência pelo valor da estimativa de ρ utilizada, sobre o qual é preciso ter bastante segurança e precisão para não levar a conclusões equivocadas.

Referências Bibliográficas

BARCIKOWSKI, R. S. (1981). Statistical Power with Group Mean as the Unit of Analysis. Journal of Educational Statistics, 6, 267-285.

BLAIR, R C., HIGGINS, J. J., TOPPING, M. E. H., & MORTIMER, A. L. (1983). An Investigation of the Robustness of the t-test to Unit of Analysis Violations.Educational and Psychological Measurement, 43, 69-80.

BLAIR, R.C., & J.J. HIGGINS(1986). Comment on statistical power with group mean as the unit of analysis. J. Edu. Stat. 2:161-169.

BOND, M., & HIGGINS, J. J. (2001). A Note on ‘A Comparison of Bayes and Maximum Likelihood Estimation of the Intraclass Correlation Coefficient’. Communications in Statistics – Theory and Methods, 30, 371-380.

FISHER, R.A. (1925). Statistical Methods for Research Workers. New York: Hafner Press.

HAMADA, M., N. BALAKRISHNAN (1998).Analysing unreplicated factorial experiments: A review with some new proposals.Stat. Sinica. 8:1–41.

MACHADO, S. & K.GIRMA (2013).Use of non-replicated observations and farm trials for guiding nutrient management decisions.Western Nutrient Management Conference. 2013. Vol. 10, 72-80.Reno, RV.

PERRETT, J.J. (2004). Using prior information on the intraclass correlation coefficient to analyze data from unreplicated and under-replicated experiments.Doctoral Dissertation, Kansas State University, Manhattan, KS.

PERRETT, J.J., & J. HIGGINS(2006). A method for analyzing unreplicated agricultural experiments. Crop Sci. 46:2482–2485.

Anexo I

Código SAS utilizado na análise do experimento:

```
/** Parte 1 – Define variáveis **/
%let p0=0.1;  /** p0 = CCI a ser utilizado **/
%let gi=1;      /** gi = número de unidades experimentais por tratamento **/
%let ti=4;      /** ti = número de tratamentos **/
/** Parte 2 – Cria a matriz que representará a variância do efeito aleatório do modelo **/
Proc iml;
RATIO=((&p0/(1-&p0))*I(&gi*&ti));     /** Cria a razão que irá compor a diagonal principal da matriz **/
create gratio from RATIO;                    /** Cria a matriz, que é a matriz identidade vezes a razão **/
append from ratio;
quit;
data gratio; set gratio;                     /** Cria o banco de dados gratio com a matriz **/
row=(_N_);
run;
/** Parte 3 - Ajusta o modelo **/
Proc mixed data=medidas ratio plots(only) = all;       /** Define o procedimento, o banco de dados e pede os
                                                             gráficos de diagnóstico do modelo **/
class metodo;                                  /** Define as variáveis categóricas **/
model pre_carga=metodo/ddfm=kr;                /** Define o modelo apropriado e os graus de liberdade ajustados
        pela razão criada. Para realizar a análise com o ajuste de Satterthwaite, trocar "ddfm=kr"  por
        "ddf = gl_satt", onde gl_satt é o número de graus de liberdade pelo ajuste, obtidos por (2.3.2) **/
random implante(metodo) / gdata=gratio Ratios;             /** Define o efeito aleatório e pede que a variância
                                                    dessa componente seja estimada pela matriz gratio criada **/
lsmeans metodo /pdiff adjust=tukey;  /** Realiza (caso aja rejeição da hipótese de igualdade de médias de
                                       tratamentos) a complementação da análise através de comparações
                                       múltiplas pelo teste de Tukey conforme (2.4.1).**/
run; quit;
```