

USANDO EXAMS NO R PARA ELABORAR QUESTIONÁRIOS AVALIATIVOS NUMA DISCIPLINA DE ESTATÍSTICA I

Naiara C. A. dos Santos^{1,2} e Jorge Luis Bazán²

¹Universidade Federal de São Carlos—UFSCar

²Universidade de São Paulo—USP

naicaroline2@usp.br

Com a eclosão da pandemia de COVID-19, o isolamento social foi imposto, e o ensino passou a ser remoto, exigindo de todos adequações e incorporação de ferramentas digitais para dar continuidade ao ano letivo e, com isso, trouxe diferentes desafios no ensino, principalmente na avaliação. Diante disso, neste trabalho apresentamos a elaboração de questionários avaliativos automatizados na área de estatística, utilizando o pacote exams no R. Os resultados mostram que as estratégias consideradas foram úteis para avaliar o desempenho da disciplina no Ensino a Distância (EaD). Adicionalmente, nossas experiências determinaram que o pacote exams fornece uma boa estrutura para geração automática de avaliações estatísticas, o que é especialmente útil para grandes quantidades de provas diferentes.

INTRODUÇÃO

Com o decreto da pandemia mundial de COVID-19 feito pela Organização Mundial de Saúde (OMS), foram necessárias várias atitudes para tentar conter a transmissão do vírus, dentre elas a quarentena e o adiamento ou cancelamento de eventos e aglomerações. As universidades, por sua vez, devido ao número grande de circulação de pessoas, foram fechadas e as aulas suspensas por tempo indeterminado (ONU, 2020). Diante disso, desde a suspensão das aulas presenciais a solução encontrada para prosseguir com o ano letivo foi o EaD.

No EaD, é necessário um ambiente próprio que permita contemplar os objetivos do curso e os anseios dos discentes. Para isso, as mais variadas ferramentas e plataformas foram e continuam sendo utilizadas pelos profissionais da educação, para que estes realizem mediações didático-pedagógicas com os discentes, facilitando assim a comunicação entre eles. O Moodle é um exemplo de ambiente virtual, projetado para fornecer a educadores, administradores e alunos um único sistema robusto e seguro para criar ambientes de aprendizagem personalizados, onde é possível a criação de repositórios de materiais para estudo, realização de atividades avaliativas, discussões em fóruns, dentre outras funcionalidades (Dougiamas et al., 2022). A escolha de plataformas pode variar de acordo com a instituição de ensino. Em particular, na nossa universidade, Universidade de São Paulo—USP/São Carlos, foi adotada a plataforma E-disciplinas, também chamada de Moodle USP, a qual se apoia no Moodle.

Os questionários Moodle são uma maneira adequada de avaliar os conhecimentos e habilidades adquiridos pelos alunos (López et al., 2010). O Moodle oferece uma ampla variedade de tipos de perguntas, desde as mais simples como resposta curta, verdadeiro ou falso, múltipla escolha ou numérica, até as mais complexas e sofisticadas, como clique e arraste textos ou imagens. A elaboração dos questionários Moodle pode ser feita de várias maneiras. Uma delas é usar os formulários fornecidos pelo próprio Moodle, mas o processo é lento e trabalhoso, e é preciso elaborar pergunta por pergunta. Se esse processo tiver que ser repetido com muitas perguntas e/ou varias versões de uma mesma pergunta, o tempo necessário acaba sendo inviável, o que acontece muito com avaliações na área de estatística. Uma alternativa é o uso de ferramentas externas ao Moodle.

Neste contexto, este trabalho tem por objetivo mostrar por meio de um breve tutorial, o processo para elaborar questionários de forma automatizada e avaliar as características psicométricas de dois questionários aplicados na disciplina de Estatística, abordando os conceitos de análise exploratória e inferência. Especificamente, apresentaremos o uso do pacote exams no R, chamado também como R-exams (Zeileis et al., 2014), o qual permite a geração automática de avaliações/questionários visando fornecer soluções adequadas para a necessidade de grande número de provas diferentes, nas quais pode-se optar por aplicar em um arquivo PDF (*Portable Document Format*) para distribuição em sala de aula, ou em XML (*eXtensible Markup Language*) para plataformas de aprendizado. Também apresentamos os resultados das análises psicométricas dos questionários aplicados usando a teoria clássica dos testes (Pasquali, 2017), o modelo de dois parâmetros da teoria de resposta ao item (TRI) (Andrade et al., 2000) e a análise da dimensionalidade dos questionários.

PACOTE R-EXAMS

A ideia básica do R-exams é ter um gerador automático de exames onde cada exercício é um arquivo independente e também fornecer diferentes mecanismos para elaborar variações aleatórias de exercícios. O pacote permite a geração automática de exames baseado em questões no formato “.Rmd” para R-Markdown (Allaire et al., 2022) ou “.Rnw” para Sweave+LaTeX (Xie, 2022), possibilitando incluir código R para geração dinâmica de elementos de questões. Um exemplo de pergunta e sua sintaxe R-Exams são apresentados na Figura 1.

Arquivo .Rnw	Arquivo .Rmd
<pre> \begin{question} Para verificar se existe correlação entre duas variáveis, selecione a(s) alternativa(s) que contém ferramenta adequada para essa análise. \begin{answerlist} \item Gráfico de Pareto. \item Boxplot. \item Coeficiente de Kurtose. \item Coeficiente de correlação. \item Gráfico de dispersão. \end{answerlist} \end{question} \exname{Questao_corr} \extype{mchoice} \exsolution{00011} </pre>	<pre> Question ===== Para verificar se existe correlação entre duas variáveis, selecione a(s) alternativa(s) que contém ferramenta adequada para essa análise. Answerlist ----- * Gráfico de Pareto. * Boxplot. * Coeficiente de Kurtose. * Coeficiente de correlação. * Gráfico de dispersão. Meta-information ===== exname: Questao_corr extype: mchoice exsolution: 00011 </pre>

Figura 1. Sintaxe do R-exams nos formatos Sweave e Markdown—Exemplo de pergunta de múltipla escolha

Os tipos de exercícios (`extype`) incluem perguntas de escolha única (`schoice`), múltipla escolha (`mchoice`), numérica (`num`), texto (`string`) ou combinações destes (`cloze`). Cada exercício (ou melhor, “modelo de exercício”) é um arquivo único com até quatro elementos: (a) Processo de geração de dados em R; (b) Texto da pergunta; (c) Texto da solução; (d) Meta-informação: Tipo de pergunta, solução correta, nome/rótulo, etc.. Apenas a questão (b) e a meta-informação (d) são obrigatórias enquanto a geração de dados (a) e a solução (c) são opcionais. As soluções podem ser usadas para informar os professores sobre a solução correta para algum exercício sorteado aleatoriamente ou para dar feedback aos alunos sobre como resolver um determinado exercício.

Com base em modelos de exercícios (potencialmente) dinâmicos, um grande número de provas, questionários ou testes personalizados pode ser criado para vários sistemas: PDFs para provas escritas clássicas, XML formatos de importação para sistemas de gerenciamento de aprendizado, como Moodle, e a possibilidade de criar saídas personalizadas como HTML, Docx, entre outros. Além disso, o R-exams oferece diferentes mecanismos para desenhar variações aleatórias de exercícios: selecionando aleatoriamente um (ou mais) de um conjunto de exercícios para cada participante; misturar aleatoriamente alternativas de resposta em perguntas de escolha única ou múltipla escolha; seleção aleatória de números, gráficos, conjuntos de dados, etc. usando a linguagem de programação R; entre outros.

O site <http://www.r-exams.org/> fornece uma visão geral, tutoriais curtos, uma galeria de modelos de exercícios e links para outros materiais. Também elaboramos tutoriais em vídeo, disponíveis em <https://jorgeluisbazan.weebly.com/presentations.html>, que mostram em etapas, de forma simples, como: (a) Instalar o software necessário para começar a gerar perguntas aleatórias; (b) Codificar perguntas e respostas; (c) Exportar essas perguntas em pdf e em XML para o Moodle.

APLICANDO O R-EXAMS PARA ELABORAR QUESTIONÁRIOS

A disciplina de Estatística I, é uma disciplina introdutória oferecida na maioria das universidades em diversos cursos. Em consequência da pandemia, essa disciplina em nossa universidade foi ofertada EaD, o que nos trouxe diferentes desafios no ensino e principalmente na avaliação. Especificamente, essa disciplina foi oferecida por meio da plataforma E-disciplinas a qual se apoia no Moodle, onde nos oferece a possibilidade de propor questionários virtuais, seja como recurso para

avaliações ou para o processo de ensino-aprendizagem. No entanto, há um grande desafio em como avaliar os conteúdos práticos da disciplina usando a modalidade EaD, uma vez que os testes virtuais feitos em casa abrem espaço para a troca de informações.

Como resposta a este problema, pode-se aumentar o número de perguntas diferentes apresentadas a diferentes alunos. Realizar esta tarefa diretamente no moodle demandaria muito trabalho e tempo, uma vez que é preciso elaborar pergunta por pergunta, se a questão tem imagens, é necessário criá-las e carregá-las para o Moodle, também é necessário configurar as notas associadas à resposta correta e às alternativas incorretas. Diante disso, o pacote de R-exams se torna uma ferramenta poderosa que fornece soluções a esses problemas: produz automaticamente questões que podem ser exportadas para o formato Moodle (XML). Desta forma, com algumas vantagens para os professores de estatística devido a familiaridade com o R, se torna simples e rápido criar um número grande de versões das perguntas e exercícios sem realmente escrevê-los tantas vezes (apenas uma vez), todas com o mesmo nível de dificuldade e relacionadas aos mesmos conteúdos.

Uma vez configuradas as perguntas e geradas as quantidades desejadas de suas versões usando a função “exams2moodle” em R (para detalhes dos argumentos dessa função consultar o *help* no R), estão prontas para que possam ser importadas no banco de questões do Moodle. Na Figura 2, apresentamos um diagrama com o resumo do processo de criação e operacionalização desses questionários.

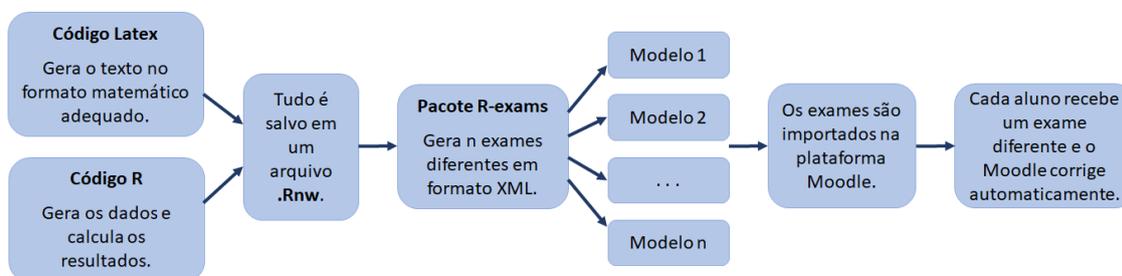


Figura 2. Resumo do processo de criação e operacionalização dos questionários

Fonte: Forner et al. (2018)

De acordo com os passos para criação de questionários usando o R-exams, elaboramos dois questionários avaliativos (Q1 e Q2, respectivamente) com cinco itens (questões) cada e cada item com cinco alternativas de respostas, das quais apenas uma era a correta (questão de escolha única, *schoice*), sendo o Q1 referente a conceitos de análise exploratória de dados e o Q2 com relação a inferência estatística. Foram gerados 50 versões dos questionários, um para cada aluno, e importados para a plataforma E-disciplinas, os quais foram submetidos como forma de avaliação para os cinquenta alunos matriculados na disciplina de Estatística I. Embora o enunciado das questões sejam os mesmos em cada pergunta, ao se usar diferentes dados cada estudante recebeu uma pergunta diferente com diferentes alternativas, tendo, portanto, gerado 50 questionários diferentes na prática. Adicionalmente, ambos questionários apresentaram aleatorização na ordem das alternativas de respostas e também aleatorizava a ordem das perguntas, com intuito de que cada estudante recebesse uma prova diferente baseado nos diferentes conjuntos de dados e nas diferentes condições usando as duas estratégias, garantindo assim que a prova pudesse maximizar a resposta mais idônea dos estudantes. Na Figura 3 podemos ver a pré-visualização de uma das questões de cada um dos questionários propostos (vista dos alunos na plataforma).

AVALIAÇÃO DAS CARACTERÍSTICAS PSICOMÉTRICAS DOS QUESTIONÁRIOS

Para realizar as análises das respostas dos estudantes aos dois questionários apresentados, usaremos a abordagem da análise clássica e a abordagem TRI (Bazán, 2018). Dentro da abordagem clássica, a primeira análise feita diz respeito à consistência interna do teste a fim de verificar a qualidade do instrumento de medida na mensuração das habilidades dos respondentes. Para isto, utilizamos o coeficiente alfa de Cronbach. Realizamos também, a análise da unidimensionalidade dos questionários por meio da análise fatorial exploratória (AFE) usando a matriz de correlação policórica. Sob o enfoque

da TRI, consideramos o modelo logístico desenvolvido por Birnbaum (1968), também conhecido como modelo de dois parâmetros (TRI-2PL), este modelo considera dois parâmetros para a análise de itens, a dificuldade e a discriminação, e é definido pelas seguintes expressões:

$$Y_{ij} | \theta_i, a_j, b_j \sim \text{Bernoulli}(p_{ij})$$

$$P(Y_{ij} = 1 | \theta_i, a_j, b_j) = \frac{e^{a_j(\theta_i - b_j)}}{1 + e^{a_j(\theta_i - b_j)}}$$

onde $i = 1, \dots, n, j = 1, \dots, k$, $P(Y_{ij} = 1 | \theta_i, a_j, b_j)$ é a probabilidade do indivíduo i com habilidade θ_i de acertar o item j , a_j é o parâmetro de discriminação do item j , com valor proporcional à inclinação da curva característica do item no ponto de abscissa b_j , que representa o parâmetro de dificuldade do item j . Todas as análises foram feitas por meio de rotinas computacionais implementadas no software R (R Development Core Team, 2022). Os questionários estudados, Q1 e Q2, possuem cinco itens cada sendo o Q1 referente a conteúdos de análise exploratória e o Q2 aborda tópicos de inferência estatística. Os resultados das análises são apresentados na Tabela 1.

Estudando-se as idades (em anos) dos funcionários de um shopping, obtiveram-se algumas medidas descritas no quadro abaixo.

Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Variância
22	29	36	37	46	50	88

Determine o coeficiente de variação (CV) e a amplitude interquartil (AI).

Escolha uma opção:

- a. CV = 24% e AI = 28
- b. CV = 2.44 e AI = 28
- c. CV = 17 e AI = 0.24
- d. CV = 0.24 e AI = 17
- e. CV = 2.38 e AI = 17

(a) Questionário 1.

Uma máquina está produzindo peças de metal com formato cilíndrico. Uma amostra é retirada e seus diâmetros (em centímetros) são:

1.3, 0.38, 1.02, 0.53, 1.37, 1.03, 0.01, 0.24, 0.37, 0.73, 0.33, 0.39, 1.6, 1.21, 0.89

Determine um intervalo de confiança de 95% para o diâmetro médio das peças dessa máquina, assumindo uma distribuição aproximadamente Normal. (Na resposta dos itens, considere duas casas decimais)

Escolha uma opção:

- a. IC95% = [0.54; 0.98].
- b. IC95% = [0.63; 0.89].
- c. IC95% = [0.65; 0.87].
- d. IC95% = [0.52; 1.00].
- e. IC95% = [0.49; 1.03].

(b) Questionário 2.

Figura 3. Exercícios baseados em R-exams—Visualização do E-disciplinas (Moodle USP)

O valor do alfa de Cronbach total para escala no modelo clássico para o Q1 foi de 0,76 considerado bom, indicando uma boa confiabilidade, por outro lado, para o Q2 foi de 0,13 considerado inaceitável. Observamos também que as médias de acerto no Q1 são maiores comparadas com as médias (Média) do Q2, mas as variabilidades (DP) são semelhantes. A correlação item-total desconsiderando o item (CIT) mostram que há uma alta correlação entre os itens dos questionários e os demais itens de Q1, já no Q2 estas correlações são baixas para os itens 1, 2 e 3 indicando que estas questões não medem o mesmo que as demais questões.

Tabela 1. Itens, estatísticas de itens baseada na TCT e resultados da AFE

Questionário 1						Questionário 2						
Análise Clássica de Itens			Análise Fatorial			Análise Clássica de Itens			Análise Fatorial			
Itens	Média	DP	CIT	Alfa	Carga Fatorial	Itens	Média	DP	CIT	Alfa	Carga Fatorial	
											F1	F2
1	0,88	0,33	0,66	0,71	0,79	1	0,74	0,44	0,11	0,17	-0,51	0,33
2	0,94	0,24	0,87	0,67	1,02	2	0,82	0,39	0,14	0,13	-0,46	0,82
3	0,82	0,39	0,63	0,70	0,84	3	0,12	0,33	0,16	0,23	1,00	0,07
4	0,86	0,35	0,63	0,71	0,84	4	0,14	0,35	0,23	0,15	0,68	0,05
5	0,80	0,40	0,44	0,77	0,61	5	0,16	0,37	0,63	-0,28	-0,26	1,00

DP: Desvio Padrão; CIT: Correlação Item-Total.

A partir dos valores observados na análise fatorial exploratória é verificada a dimensionalidade da escala. Considerando todos os itens há um único fator para o Q1 e dois fatores para o Q2. Os gráficos na Figura 4, confirmam um modelo unidimensional e bidimensional para os questionários Q1 e Q2 respectivamente. O desempenho ruim do Q2 pode ser explicado as baixas taxas de acerto observadas nos itens 3, 4 e 5, que correspondem, respectivamente, a teste de hipóteses para média de uma população, para média de duas populações com variâncias desconhecidas e iguais e, para média de duas populações com variâncias desconhecidas e diferentes.

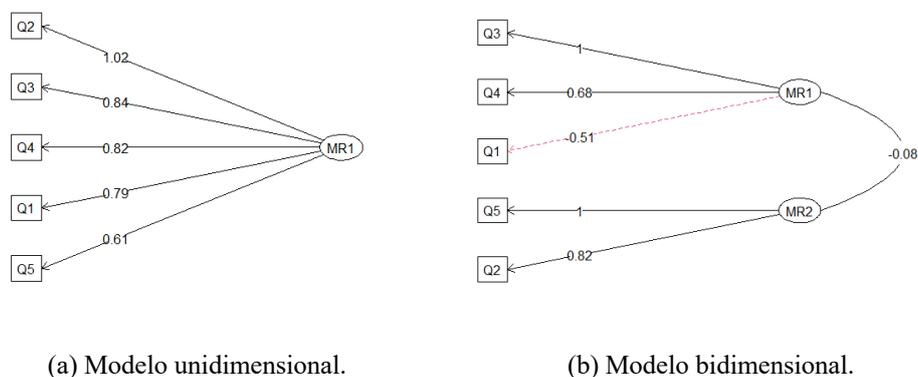


Figura 4. Análise fatorial por teoria clássica dos itens

Com base nesses resultados de unidimensionalidade, determinamos que o Q1 de 5 itens é suficientemente unidimensional para a análise de TRI. Embora os resultados para o Q2 mostraram bidimensionalidade nos ajustamos o modelo TRI para ele. Os resultados da análise TRI-2PL para os questionários estão apresentados na Tabela 2. Apesar de não haver um valor consensual ou um critério de corte para o parâmetro de discriminação a , Hafsteinsson et al. (2007) sugerem que quando existem poucos itens em uma escala (valores inferiores a 10 itens), para que se tenha um padrão elevado de qualidade do item, e, portanto, da qualidade da medição, pode-se requerer um valor de 2,0 ou superior.

Tabela 2. Análise TRI-2PL para os itens dos questionários

Questionário 1			Questionário 2		
Item	Parâmetros		Item	Parâmetros	
	b	a		b	a
1	-1,4599	2,3229	1	-1,0180	1,3837
2	-1,5778	31,4342	2	-1,3301	1,6390
3	-1,0592	2,7435	3	-1,3337	-3,3835
4	-1,3434	2,2870	4	-1,5726	-1,6448
5	-1,3068	1,4186	5	26,0283	0,0638

No caso do questionário 1, exceto o item 5 todos os outros apresentaram valores superiores a 2, atingindo assim valores aceitáveis. Ainda no Q1, podemos observar que o item 2 apresenta um alto valor de discriminação $a = 31,43$. Já para o questionário 2, todos os valores são considerados não aceitáveis. Para o parâmetro de dificuldade do item, o espaço paramétrico para b_j é arbitrário e corresponde ao mesmo que o traço latente, o qual esta geralmente na reta real. O ideal em uma avaliação é observarmos itens fáceis, médios e difíceis, ou seja, valores de b_j negativos, zeros e positivos. No entanto, não é o que acontece para ambos questionários, em que todos itens apresentam valores negativos indicando questões muito fáceis, exceto para o item 5 do Q2, o qual apresenta um valor de $b = 26,03$ indicando ser um item muito difícil em relação aos demais. Este item referente ao conteúdo de teste de hipóteses para média de duas populações com variâncias desconhecidas e diferentes.

CONCLUSÕES

Frente a essa nova realidade da educação, com a chegada da pandemia do COVID-19, em que as escolas e universidades tiveram que aderir o ensino remoto enquanto é vigorado o isolamento social, os professores apresentam muitos desafios para ministrar suas aulas e avaliar os alunos no novo formato, tendo que usar novas metodologias e fazendo uso de plataformas digitais. Os questionários do Moodle são uma maneira conveniente de lidar com a avaliação na modalidade EaD, no entanto, sua elaboração é trabalhosa e lenta. Diante disso, apresentamos uma ferramenta para auxiliar os professores no processo de elaboração desses questionários avaliativos, permitindo a geração automática de questões. Adicionalmente, nossas experiências determinaram que o pacote R-exams fornece uma estrutura para geração automática de exames estatísticos, o que é especialmente útil quando é necessário um grande número de exames diferentes. O principal impacto da utilização desta ferramenta está na enorme redução do tempo gasto na elaboração de questões, substituindo muito trabalho com formulários Moodle por poucos passos. Assim, os usuários não devem se preocupar com detalhes de implementação e podem se concentrar na especificação do conteúdo e no desenvolvimento do conjunto de exercícios. Enquanto a estrutura do pacote pode ajudar no desenho e desenvolvimento dos exercícios, o pacote não pode, é claro, garantir que “bons” exercícios (do ponto de vista educacional) sejam gerados, como podemos observar com as análises realizadas, onde o Q1 apresentou uma boa confiabilidade e o Q2 não, e ambos questionários apresentaram itens muito fáceis, entretanto, estes podem ser aprimorados em propostas futuras.

REFERÊNCIAS

- Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., & Iannone, R. (2022). *Rmarkdown: Dynamic documents for R* (R package version 2.14) [Computer software]. <https://rmarkdown.rstudio.com>
- Andrade, D. F., Tavares, H. R., & da Cunha Valle, R. (2000). Teoria da resposta ao Item: Conceitos e aplicações. *Simpósio Nacional de Probabilidade e Estatística*. Associação Brasileira de Estatística.
- Bazán, J. L. (2018). Psicometria e avaliação por testes: Um marco metodológico. *Avaliação da educação: Referências para uma primeira conversa*. EdUFSCar.
- Birnbaum, A. L. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–479). Addison-Wesley
- Dougiamas, M., Moodle HQ, & Moodle Community. (2022). *Moodle* (Versão 4.0) [Computer software]. Moodle Project. <https://moodle.org/>
- Forner, Ò., Gandía, C., Molina, M. D., Mulero, J., Nueda, M. J., & Pascual, A. (2018). REDEST—Cuestionarios aleatorios en Moodle. En R. Roig-Vila, (Coord.), J. M. A. Martinez, A. L. Carreres, & N. P. Buades (Eds.), *Memorias del Programa de Redes-I3CE de calidad, innovación e investigación en docencia universitaria. Convocatoria 2017-18*. Universitate D'Alacant.
- López, J. M., Romero, E., & Ropero, E. (2010). Utilización de Moodle para el desarrollo y evaluación de competencias en los alumnos. *Formación Universitaria*, 3(3), 45–52. <https://doi.org/10.4067/S0718-50062010000300006>
- Organização das Nações Unidas. (2020, March 10). *COVID-19: Unesco divulga 10 recomendações sobre ensino a distância devido ao novo coronavírus*. ONU News. <https://news.un.org/pt/story/2020/03/1706691>
- Pasquali, L. (2017). *Psicometria: Teoria dos testes na psicologia e na educação*. Editora Vozes.
- R Development Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Xie, Y. (2022). *Tinytex: Helper functions to install and maintain TeX live, and compile LaTeX documents* (R package version 0.39) [Computer software]. <https://github.com/rstudio/tinytex>
- Zeileis, A., Umlauf, N., & Leisch, F. (2014). Flexible generation of e-learning exams in R: Moodle quizzes, OLAT assessments, and beyond. *Journal of Statistical Software*, 58(1), 1–36. <https://doi.org/10.18637/jss.v058.i01>