



## TEORIA CLÁSSICA DOS TESTES EM AMBIENTE R COMO APOIO NO ENSINO DE ENGENHARIA

**Olga Maria Formigoni Carvalho Walter** – [olga@hotmail.com.br](mailto:olga@hotmail.com.br)

Universidade Federal de Santa Catarina, Depto.de Engenharia de Produção e Sistemas  
Campus Universitário Trindade. Caixa Postal 476  
CEP 88040-900 – Florianópolis/SC - Brasil.

**Andréa Cristina Konrath** – [andreack@inf.ufsc.br](mailto:andreack@inf.ufsc.br)

Universidade Federal de Santa Catarina, Departamento de Informática e Estatística  
Campus Universitário Trindade. Caixa Postal 476  
CEP 88040-900 – Florianópolis/SC - Brasil

**Elisa Henning** – [elisa.henning@udesc.br](mailto:elisa.henning@udesc.br)

Universidade do Estado de Santa Catarina, Departamento de Matemática  
Rua Paulo Malschitzki, s/n. Campus Universitário Prof. Avelino Marcante. Zona Industrial  
CEP 89219-710 – Joinville/SC - Brasil.

**Antonio Cezar Bornia** – [cezar.bornia@ufsc.br](mailto:cezar.bornia@ufsc.br)

Universidade Federal de Santa Catarina, Depto.de Engenharia de Produção e Sistemas  
Campus Universitário Trindade. Caixa Postal 476  
CEP 88040-900 – Florianópolis/SC - Brasil.

**Resumo:** *Este artigo é um estudo exploratório e tem como objetivo explorar alguns pacotes do R específicos da Teoria Clássica dos Testes (TCT) como o ltm e CTT apresentando aplicações com dados do Exame Nacional de Desempenho de Estudantes (ENADE) de 2014 do curso de Engenharia de Produção. O R, por se tratar de um software livre e por empregar linguagem de programação, permite desenvolver suas próprias funções. Além disso, possui uma ampla variedade de métodos estatísticos já implementados que são úteis em análises de testes. Os resultados demonstram que ambos os pacotes ltm e CTT demonstram-se estáveis, podendo, assim, serem utilizados como meio de ampliar os recursos utilizados em sala de aula para auxiliar no processo de ensino aprendizagem. Conclui-se que a confiabilidade do conjunto de dados analisado está no limiar do que é recomendado pela literatura e os itens têm pouco poder de discriminação. Além disso, os itens que possuem maior índice dificuldade são os do conteúdo profissionalizante. Este artigo visa contribuir para a disseminação da TCT em termos de linguagem R, permitindo que tanto educadores quanto estudantes utilizem as informações aqui dispostas em suas análises de dados.*

**Palavras-chave:** *Teoria Clássica dos Testes, software R, ENADE, Engenharia de Produção.*

Organização



**UDESC**  
UNIVERSIDADE  
DO ESTADO DE  
SANTA CATARINA



Promoção





## 1 INTRODUÇÃO

O crescente avanço e disseminação de tecnologias de informação e comunicação vem causando transformações em todas as áreas do conhecimento e afetando os mais diversos ambientes de nossa sociedade, seja ele cultural, político, econômico e educacional. Na atualidade, as relações do homem com o mundo estão crescentemente sendo mediadas por recursos tecnológicos. De acordo com Leite (2014), a introdução de novas tecnologias no ensino produz resultados positivos na aprendizagem, pois modifica e melhora a maneira como os professores ensinam, além da forma como os alunos aprendem.

Para otimizar o aprendizado em sala de aula, é necessário avaliar periodicamente o desempenho do aluno, bem como a maneira como o conhecimento é transmitido. Essa prática permite o aperfeiçoamento do processo de ensino aprendizagem. Nesse sentido, o meio de avaliação (teste, prova, exame) deve ser bem elaborado. Conforme Coelho (2014), a avaliação do aprendizado, da aquisição de habilidades e das competências alcançadas pelos estudantes de nível superior tem sido realizada de maneira multidisciplinar pelas instituições de ensino superior. Os efeitos dessa avaliação geram informações capazes de analisar o perfil dos cursos de graduação, colaborando para o avanço e melhoria da qualidade do ensino aprendizagem.

A maneira mais usual para avaliação de desempenho educacional é a Teoria Clássica de Testes (TCT) (SARTES; FORMIGONI, 2013), que se baseia na quantidade de questões corretas, para obter o *score* total do teste. Uma vantagem da TCT é que ela é relativamente fácil de interpretar e exige poucas suposições sobre os dados (JUSTINO; ANDRADE, 2007).

De acordo com o exposto, esse artigo tem como objetivo explorar alguns pacotes do R, específicos da TCT como o ltm (RIZOPOULOS, 2006) e CTT (WILLSE; SHU, 2008) apresentando algumas aplicações de avaliação de curso de nível superior.

## 2 TEORIA CLÁSSICA DOS TESTES

A TCT é composta por conceitos e técnicas que têm sido utilizados como base para o desenvolvimento de vários instrumentos de avaliação e como ponto de partida e de referência para abordagens de medidas psicométricas (DEVELLIS, 2006) podendo ser utilizada numa grande variedade de situações.

Para Sartes e Formigoni (2013), há alguns fatores que explicam a ampla utilização da TCT: i) a maioria dos cursos de graduação e de pós-graduação foca na TCT e abordam pouco sobre outros métodos estatísticos de avaliação de questionários; ii) a TCT é mais simples de entender do que outras metodologias de avaliação e; iii) o tamanho da amostra para aplicação da TCT não exige ser muito grande.

A TCT utiliza o *score* no teste como sua referência de medida e tem como objetivo representar com números (medidas) as pessoas e os itens a partir das respostas. O modelo da TCT é dado por  $X=V+E$ , que indica que o *score* observado, ou seja, a soma de pontos no teste ( $X$ ) é resultante da soma do valor verdadeiro ( $V$ ) - valor autêntico se não houvesse erro na medição - com o erro aleatório de medição ( $E$ ). Assim,  $X$  é uma estimativa do *score* verdadeiro.

Analogamente, pode-se assumir que a variância total dos valores observados  $\sigma_X^2$  é composta pela soma das variâncias dos valores verdadeiros  $\sigma_V^2$  e a variância dos erros  $\sigma_E^2$  conforme Equação (1) (HAYES, 1995; HORA; MONTEIRO; ARICA, 2010):

$$\sigma_X^2 = \sigma_V^2 + \sigma_E^2 \quad (1)$$

Organização



**UDESC**  
UNIVERSIDADE  
DO ESTADO DE  
SANTA CATARINA



Promoção





Algumas medidas utilizadas na TCT são os índices de dificuldade e discriminação; os coeficientes de correlação bisserial e ponto bisserial e o coeficiente Alpha de Cronbach. O índice de dificuldade varia entre 0 e 1 e é dado pela proporção de respostas corretas em cada item, para itens dicotômicos. Para itens politômicos, o índice de dificuldade é calculado pela proporção de respostas de uma categoria de escolha ou pela média das respostas de todos os respondentes (GRÉGOIRE; LAVEAULT, 2002; DEVELLIS, 2006). Quanto mais próximo de um for o índice de dificuldade, mais difícil é o item.

O índice de discriminação varia entre -1 e 1, sendo mais discriminativo quanto maior for o seu valor. Tem o objetivo de diferenciar o grupo de indivíduos que tiveram *scores* mais altos dos que tiveram *scores* mais baixos no teste. É dado pela diferença entre a proporção de acertos dos participantes com maior habilidade (27% daqueles com pontuações mais altas) daqueles com menor habilidade (27% daqueles com pontuações mais baixas) (KELLEY, 1939).

O coeficiente ponto bisserial é definido como a correlação de Pearson entre uma variável dicotômica e o *score* do teste (desempenho no teste) de acordo com a Equação (2) (ANJOS; ANDRADE, 2012):

$$\rho_{pb} = \frac{\bar{X}_A - \bar{X}_T}{S_T} \sqrt{\frac{p}{(1-p)}}; \quad (2)$$

em que,  $\bar{X}_A$  é a média dos *scores* dos respondentes que acertaram o item;  $\bar{X}_T$  é a média global dos *scores* do teste;  $S_T$  é o desvio padrão do teste e  $p$  é a proporção de respondentes que acertaram o item.

O coeficiente bisserial é uma transformação no coeficiente ponto bisserial. Neste caso, o desempenho no item dicotomizado, tem distribuição Normal. É uma medida de associação entre uma variável dicotomizada e uma variável contínua, e é definido pela Equação (3) (ANJOS; ANDRADE, 2012):

$$\rho_b = \rho_{pb} \frac{\sqrt{p(1-p)}}{h(p)}; \quad (3)$$

em que,  $\rho_{pb}$  é a correlação ponto bisserial;  $p$  é a proporção de respondentes que acertaram o item e  $h(p)$  é o valor da densidade da distribuição normal padrão no ponto em que a área da curva à esquerda deste ponto é igual a  $p$ .

O coeficiente alfa de Cronbach (CRONBACH, 1951) é a estatística mais comumente aplicada para medir a consistência interna de um teste (VIEIRA, 2017). Consistência interna de um teste é a extensão em que os itens que compõem o teste medem o mesmo conceito ou construto.. Seja  $x_{ij}$  o  $i$ -ésimo *score* do  $j$ -ésimo respondente,  $i = 1, 2, \dots, k$ , e  $j = 1, 2, \dots, n$ . Considere um questionário com  $k$  itens, respondido por  $n$  pessoas. Para calcular o coeficiente alfa de Cronbach ( $\alpha$ ), aplica-se a Equação (4):

$$\alpha = \frac{k}{k-1} \left( 1 - \frac{\sum_{i=1}^k S_i^2}{S_T^2} \right); \quad (4)$$

em que,  $k$  é o número de itens;  $\sum_{i=1}^k S_i^2$  é a soma das variâncias dos  $k$  itens;  $\bar{X}_T$  é a média global dos *scores* do teste e  $S_T^2$  é a variância global dos *scores* do teste.

O coeficiente  $\alpha$  varia entre 0 a 1 e quanto mais próximo de 1 maior a consistência do teste (ANJOS; ANDRADE, 2012). Há discussão sobre os valores aceitáveis de alfa de Cronbach, em geral, variam entre 0,70 a 0,95. A maneira prática de julgar o valor de  $\alpha$  é comparar o valor calculado com o valor preconizado por diferentes autores em tabelas apresentadas na literatura (VIEIRA, 2017). Se as variáveis (itens) são positivamente correlacionadas, a variância da soma aumentará. Se os itens de uma escala são todos idênticos, a correlação é



perfeita, todas as  $S_i^2$  serão iguais e  $\alpha = 1$ . Por outro lado, se os itens são todos independentes,  $\alpha = 0$  (MATTHIENSEN, 2011).

### 3 MATERIAIS E MÉTODOS

Este estudo é exploratório, uma vez que investiga a utilização de pacotes do *software* R para análise da TCT. Os dados utilizados estão disponíveis no Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP, 2017). Tratam-se das respostas do ENADE de 2014, que foi aplicado a 17.307 estudantes concluintes de Engenharia de Produção. As observações de 2.225 alunos ausentes, respostas duplicadas ou em branco foram desconsideradas das análises.

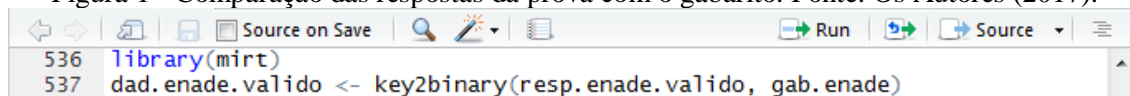
O caderno de provas em análise é composto por três partes. A primeira é a de formação geral, contendo duas questões discursivas e oito questões objetivas (item 1 até item 8); a segunda parte, de formação específica, é composta por três questões discursivas, por vinte sete questões objetivas que englobam conteúdos básicos (dez questões – item 9 até item 18) e conteúdos profissionalizantes (dezessete questões – item 19 até item 35). A última parte é composta pelo questionário de percepção da prova e contém dez questões.

Os itens que serão analisados são as questões objetivas de formação geral e específica, com exceção do item 26, do conteúdo profissionalizante que foi anulado. A análise de dados foi realizada com o R (R CORE TEAM, 2017) utilizando os pacotes ltm (RIZOPOULOS, 2006) e CTT (WILLSE; SHU, 2008).

### 4 APLICAÇÃO DA TEORIA CLÁSSICA DOS TESTES COM R

Inicialmente, os microdados do ENADE de 2014 foram dicotomizados, ou seja, como as respostas são apresentadas na forma de alternativas A, B, C, D e E estas foram comparadas ao gabarito. A Figura 1 apresenta o processo de comparação das respostas dos respondentes com o gabarito por meio da função “`key2binary( )`”. Para utilizar a função é necessário ativar o pacote “`mirt`”, por meio da função “`library( )`”.

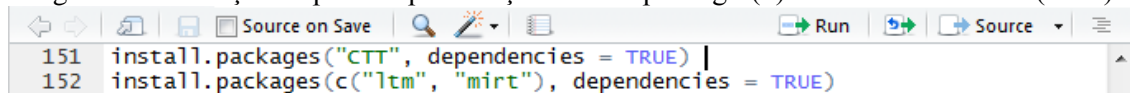
Figura 1 - Comparação das respostas da prova com o gabarito. Fonte: Os Autores (2017).



```
536 library(mirt)
537 dad.enade.valido <- key2binary(resp.enade.valido, gab.enade)
```

O *download* e instalação de pacotes no R pode ser feita pela função “`install.packages( )`”. A primeira linha da Figura 2 apresenta o processo de instalação do pacote CTT.

Figura 2 - Instalação de pacotes pela função “`install.packages( )`”. Fonte: Os Autores (2017).



```
151 install.packages("CTT", dependencies = TRUE) |
152 install.packages(c("ltm", "mirt"), dependencies = TRUE)
```

O argumento “`dependencies = TRUE`” instala automaticamente todos os outros pacotes dos quais o pacote “CTT” depende. Para instalar mais de um pacote simultaneamente, o vetor “`c( )`” pode ser utilizado para especificar a sequência de pacotes desejada, como na segunda linha da Figura 2, em que foi feita a instalação dos pacotes “`ltm`” e “`mirt`”.



#### 4.1 Análise pela Teoria Clássica dos Testes com o pacote CTT

A ativação do pacote também pode ser feita pela função “require( )” (Figura 3).

Figura 3 - Ativação do pacote CTT e frequência das respostas. Fonte: Os Autores (2017).

```
Console C:/
> require(CTT)
> apply(dad.enade.valido, 2, table)
  i1  i2  i3  i4  i5  i6  i7  i8  i9  i10 i11 i12 i13 i14  i15  i16  i17
0 6295 7594 4112 2162 8802 5752 5766 2568 9513 12351 6675 9894 5734 8903 11709 9163 5178
1 8121 6822 10304 12254 5614 8664 8650 11848 4903 2065 7741 4522 8682 5513 2707 5253 9238
  i18 i19 i20 i21 i22 i23 i24 i25 i27 i28 i29 i30 i31 i32 i33 i34 i35
0 2192 5999 10105 4790 10075 9659 12775 6127 13391 3878 9290 8183 9867 7504 8114 12488 12003
1 12224 8417 4311 9626 4341 4757 1641 8289 1025 10538 5126 6233 4549 6912 6302 1928 2413
```

A função “apply( )” (Figura 3), retorna a quantidade de respostas de cada alternativa do objeto “dad.enad”. Observa-se que os itens 24 e 27 foram os que tiveram as menores quantidades de acerto, com 1641 e 1025 acertos, respectivamente. Já os itens 4 e 18 foram os mais acertados, com 12254 e 12224 acertos, respectivamente. Os itens 24 e 27 são de conteúdo específico profissionalizante abordando pesquisa operacional e custos, respectivamente. O item 4 é de formação geral e aborda o uso de redes sociais. O item 18, relativo a materiais, é do núcleo específico de conteúdo básico.

No pacote “CTT”, a análise de confiabilidade pode ser obtida pela função “reliability( )” (Figura 4). O alfa de Cronbach resultou em 0,676. Por meio da função “names( )” é possível verificar as estatísticas calculadas pela função “reliability( )”.

Figura 4 - Alfa de Cronbach e estatísticas da função “reliability( )”. Fonte: Os autores (2017).

```
Console C:/
> confiabilidade.CTT<-reliability(dad.enade.valido, itemal=TRUE)
> confiabilidade.CTT
Number of Items      Number of Examinees      Coefficient Alpha
34                   14416                      0.676

> names(confiabilidade.CTT)
[1] "nItem"      "nPerson"    "alpha"      "scaleMean"  "scaleSD"
[6] "alphaIfDeleted" "pBis"      "bis"        "itemMean"
```

Na função “reliability( )”, o argumento “itemal = TRUE” fornecerá a média de cada item, ou seja, a soma de todas as respostas dividida pela quantidade de itens (“\$itemMean”); a correlação ponto bisserial (“\$pBis”), referindo-se ao valor com a exclusão do item, na ordem em que é apresentado; a correlação bisserial (“\$bis”) e o alfa de Cronbach excluindo o item correspondente (“\$alphaIfDeleted”).

A correlação ponto bisserial e a correlação bisserial são apresentadas na Figura 5. Os itens 10, 12 e 35 possuem os valores de correlação ponto bisserial e correlação bisserial muito baixos. Os itens 27 e 34 apresentaram tanto correlação ponto bisserial quanto correlação bisserial negativos e devem ser retirados das análises futuras por outras técnicas, como por exemplo, pela Teoria da Resposta ao Item (TRI).

Demais itens, na sua maioria, também apresentam correlação ponto bisserial e correlação bisserial baixos. Valores baixos de correlação ponto bisserial indicam que o item não apresenta boa correlação com o *score* bruto produzido pelo conjunto das variáveis. Não apresenta assim uma boa discriminação para o índice que deseja produzir. Isto significa que alunos com *score* mais alto erraram estas questões.



Figura 5 - Correlação ponto biserial e correlação biserial. Fonte: Os Autores (2017).

```

Console C:/Olga/Dout 2016/Artigos 2017/TRI com R - COBENGE/Script/
> confiabilidade.CTT$pBis
[1] 0.17442700 0.11308265 0.32612977 0.24766282 0.11832838 0.20341913
[7] 0.28319630 0.15538408 0.23063028 0.02360992 0.31936619 0.02173375
[13] 0.31742889 0.22245884 0.10029007 0.13328318 0.32507241 0.22107207
[19] 0.37551933 0.20204621 0.29238478 0.08301177 0.19574680 0.13102665
[25] 0.34848260 -0.05264605 0.28605782 0.20638912 0.21323051 0.22395104
[31] 0.27142554 0.21585755 -0.07895711 0.03332105
> confiabilidade.CTT$bBis
[1] 0.22030584 0.14184432 0.44399552 0.39227475 0.14964180 0.25840467
[7] 0.36088881 0.23020261 0.29205473 0.03658831 0.40053148 0.02790050
[13] 0.40380415 0.27919291 0.14204790 0.16855755 0.42045281 0.34687826
[19] 0.47524228 0.26086764 0.38519167 0.10846418 0.24953194 0.20167801
[25] 0.43986005 -0.09705847 0.39060507 0.26149014 0.26665442 0.28527782
[31] 0.33848231 0.27021462 -0.12535978 0.04809054
    
```

A prova analisada apresenta apenas seis itens com correlação ponto biserial maior que 0,30 (itens 3, 11, 13, 17, 19 e 25).

O *score* de cada respondente é obtido por meio da função “score( )”, conforme apresenta a Figura 6 que contém os *scores* dos primeiros quinze respondentes.

Figura 6 - *Score* dos trinta primeiros respondentes. Fonte: Os Autores (2017).

```

Console C:/
> score(resp.enade.valido[1:15,],gab.enade)
$score
  P1  P2  P3  P4  P5  P6  P7  P8  P9  P10 P11 P12 P13 P14 P15
 24  23  18  20  17  26  26  26  22  24  16  19  23  19  24
    
```

O pacote CTT também permite realizar a análise dos distratores (alternativas de respostas diferentes da alternativa correta) pela função “distractor.analysis( )”. A Figura 7 apresenta alguns itens do instrumento analisados. A coluna *lower*, *middle* e *upper* representam a quantidade de alunos com pior desempenho, desempenho médio e melhor desempenho, respectivamente. Espera-se que os alunos classificados como *lower* tenham uma maior quantidade de respostas incorretas e alunos classificados como *upper* escolham uma maior quantidade de respostas corretas.

Figura 7 - Análise dos distratores. Fonte: Os Autores (2017).

```

Console C:/
> avaliacao.distratores<-distractor.analysis(resp.enade.valido,gab.enade)
> avaliacao.distratores$i10
score.level
response lower middle upper
  A    362    254    189
  B   1600   1448   1406
  C   2034   1884   1722
 *D    524    720    821
  E    562    496    394
> avaliacao.distratores$i12
score.level
response lower middle upper
 *A   1386   1409   1727
  B    729    571    446
  C    599    549    605
  D   1731   1555   1112
  E    637    718    642
> avaliacao.distratores$i27
score.level
response lower middle upper
  A    488    359    337
  B    465    271    150
  C    485    319    313
  D   3254   3527   3423
 *E    390    326    309
    
```

De acordo com Haladyna (1994), bons distratores atraem a mesma quantidade de respondentes e distratores com pouca quantidade de respostas devem ser revisados. Se um distrator atrai mais respondentes com *scores* maiores do que a resposta correta, este também deve ser revisado. Um bom distrator não deve dar chance ao aluno que sabe ficar em dúvida com a resposta (RODRIGUES, 2006).



Os itens 10, 12, 15, 22, 24, 27, 34 e 35 apresentam este tipo de problema. Como exemplo, tem-se o item 10 (Figura 7) que contém dois distratores (B e C) com maior quantidade de respostas /do que a alternativa correta (alternativa D).

O item 27, além disso, apresenta uma situação ainda pior: a alternativa correta foi selecionada por mais respondentes de menor *score* (390) do que por alunos de maior *score* (309). Isto pode indicar que as alternativas podem ter sido mal elaboradoras.

## 4.2 Análise pela Teoria Clássica dos Testes com o pacote ltm

No pacote ltm, a correlação ponto biserial (a correlação biserial entre o *score* e um item) pode ser obtida por meio da função “`biserial.cor( )`”, conforme Figura 8, onde é apresentada a informação para o primeiro item. Esta função considera todos os itens para efetuar o cálculo da correlação ponto biserial.

Figura 8 - Correlação ponto biserial do item 1 com o pacote ltm. Fonte: Os Autores (2017)

```
Console C:/Olga/Dout 2016/Artigos 2017/TRI com R - COBENGE/Script/
> biserial.cor(rowSums(dad.enade.valido), dad.enade.valido[,1])
[1] -0.2809036
> biserial.cor(rowSums(dad.enade.valido), dad.enade.valido[,1],level=2)
[1] 0.2809036
```

A função “`biserial.cor( )`” utiliza como padrão o valor 0 (zero), ou seja, a função considera o primeiro nível das respostas para obter o valor da correlação. Para utilizar o valor 1 utilize o argumento “`level=2`”. A função “`descript( )`” apresenta uma série de estatísticas descritivas que são utilizadas na análise das respostas de questionários que são discutidas nas Figuras 9 a 11. Na Figura 9, na primeira coluna está a proporção de erros e na segunda a proporção de acertos de cada item.

Figura 9 - Proporção de erro e índice de dificuldade dos itens. Fonte: Os Autores (2017).

```
Console C:/
> descript(dad.enade.valido)

Descriptive statistics for the 'dad.enade.valido' data-set

Sample:
 34 items and 14416 sample units; 0 missing values

Proportions for each level of response:

      logit      logit      logit
i1  0.4367 0.5633 0.2547  i13 0.3978 0.6022 0.4148  i25 0.4250 0.5750 0.3022
i2  0.5268 0.4732 -0.1072  i14 0.6176 0.3824 -0.4793  i27 0.9289 0.0711 -2.5699
i3  0.2852 0.7148 0.9186   i15 0.8122 0.1878 -1.4645  i28 0.2690 0.7310 0.9997
i4  0.1500 0.8500 1.7348   i16 0.6356 0.3644 -0.5564  i29 0.6444 0.3556 -0.5946
i5  0.6106 0.3894 -0.4497  i17 0.3592 0.6408 0.5789  i30 0.5676 0.4324 -0.2722
i6  0.3990 0.6010 0.4096   i18 0.1521 0.8479 1.7186  i31 0.6844 0.3156 -0.7743
i7  0.4000 0.6000 0.4056   i19 0.4161 0.5839 0.3387  i32 0.5205 0.4795 -0.0822
i8  0.1781 0.8219 1.5290   i20 0.7010 0.2990 -0.8519  i33 0.5628 0.4372 -0.2527
i9  0.6599 0.3401 -0.6628  i21 0.3323 0.6677 0.6979  i34 0.8663 0.1337 -1.8683
i10 0.8568 0.1432 -1.7886  i22 0.6989 0.3011 -0.8420  i35 0.8326 0.1674 -1.6043
i11 0.4630 0.5370 0.1482  i23 0.6700 0.3300 -0.7083
i12 0.6863 0.3137 -0.7830  i24 0.8862 0.1138 -2.0522
```

Os dois itens mais fáceis (Figura 9) são o 4 e 18 com 85,00% e 84,79% de proporção de acertos, respectivamente. Já os dois itens mais difíceis são o 24 e o 27 com 11,38% e 7,11% de proporção de acertos, respectivamente. A função “`descript( )`” também apresenta a frequências dos *scores* (Figura 10).

Organização

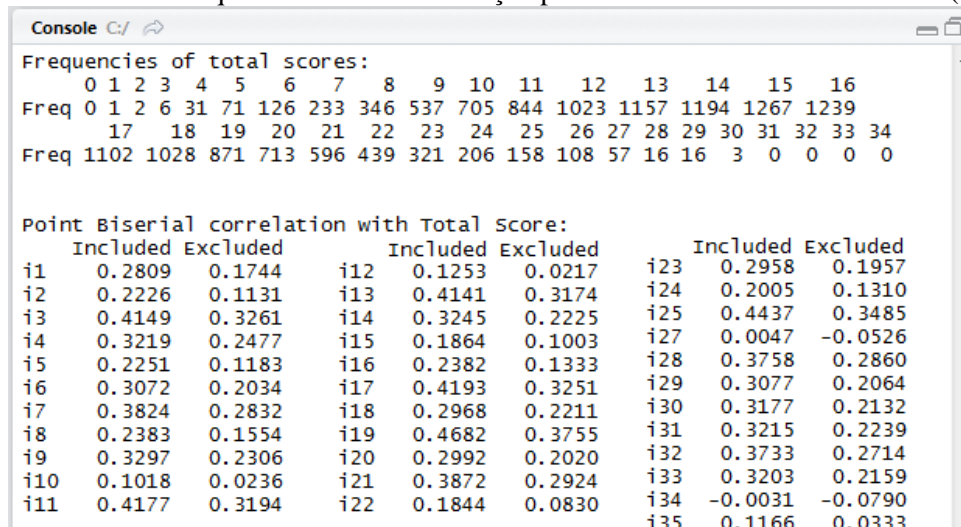


Promoção





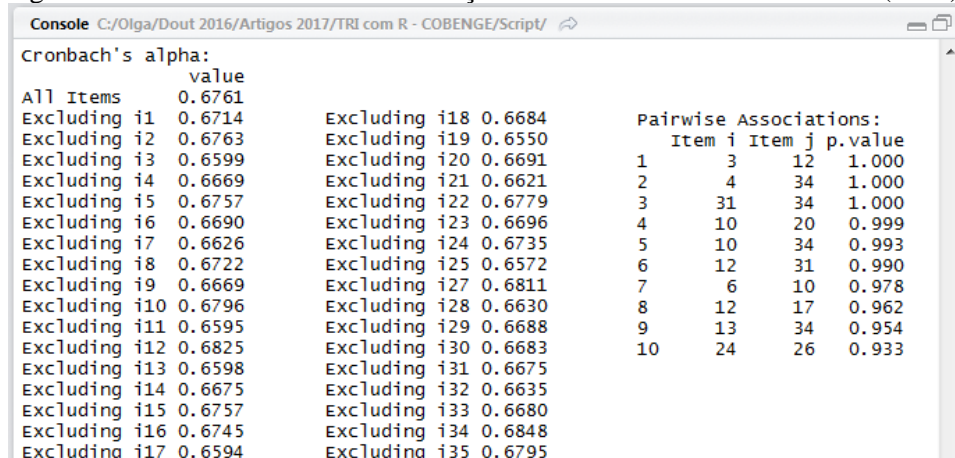
Figura 10 - Fonte: Frequência *scores* e correlação ponto biserial. Fonte: Os Autores (2017).



Observa-se, na Figura 10, que setecentos e cinco estudantes acertaram dez questões. E nenhum estudante acertou todas as questões (frequência do *score* 34 está zerada). Apenas três estudantes acertaram 30 questões na prova. Observam-se também a correlação ponto biserial com todos os itens e excluindo-se o item correspondente.

Na Figura 11, é apresentado  $\alpha = 0,6761$  para todos os itens e também  $\alpha$  excluindo-se o item correspondente. Este último mostra o impacto que teria a retirada de cada item considerado sobre o desempenho deste indicador, o que representa uma medida de contribuição de cada item para a confiabilidade do questionário.

Figura 11 - Alfa de Cronbach e associações entre itens. Fonte: Os Autores (2017).



Se  $\alpha$  aumenta com a exclusão do item deve-se considerar a possibilidade de excluí-lo (PEREIRA, 2001). A confiabilidade do teste pode ser considerada baixa. Contudo, em alguns cenários de investigação das ciências sociais,  $\alpha = 0,60$  é considerado aceitável desde que os resultados obtidos com o instrumento sejam interpretados com precaução (DEVELLIS, 1991). Na Figura 11, são apresentadas as associações entre os itens. Associações não significativas podem indicar itens com problemas (DEVELLIS, 2006).





O pacote ltm dispõe da função “cronbach.alpha( )” (Figura 12) para cálculo de  $\alpha$  de todos os itens do instrumento. O argumento “CI = TRUE” apresenta o intervalo de confiança para a estimativa do alfa de Cronbach.

Figura 12 - Função do alfa de Cronbach do pacote ltm. Fonte: Os Autores (2017).

```
Console C:/Olga/Dout 2016/Artigos 2017/TRI com R - COBENGE/Script/
> cronbach.alpha(dad.enade.valido, CI = TRUE)

Cronbach's alpha for the 'dad.enade.valido' data-set

Items: 34
Sample units: 14416
alpha: 0.676

Bootstrap 95% CI based on 1000 samples
2.5% 97.5%
0.669 0.683
```

## 5 CONCLUSÃO E CONSIDERAÇÕES FINAIS

Este artigo explorou os pacotes ltm e CTT do R, na TCT com aplicações do ENADE de 2014 do curso de Engenharia de Produção. Ambos os pacotes utilizados demonstram-se estáveis e confiáveis, podendo assim, serem utilizados como meio de ampliar os recursos utilizados para auxiliar no processo de ensino aprendizagem.

A confiabilidade do conjunto de dados está no limiar do que é recomendado pela literatura. Sugere-se uma análise mais aprofundada, para avaliar a dimensionalidade do instrumento, e verificar o comportamento do alfa de Cronbach, já que a prova é composta de três construtos teóricos: formação geral, conteúdos básicos e conteúdos profissionalizantes. Foi verificado que os itens têm pouco poder de discriminação e vários itens apresentam problemas com os distratores. Os itens com maior índice dificuldade são de formação profissionalizante.

Como sugestão para trabalhos futuros, sugere-se a continuidade da análise, com outras técnicas, como a Teoria da Resposta ao Item e também, analisar separadamente os resultados das instituições públicas e privadas.

## REFERÊNCIAS BIBLIOGRÁFICAS

ANJOS, A.; ANDRADE, D. F. **Teoria da Resposta ao Item com uso do R**. ABE - Associação Brasileira de Estatística. João Pessoa, 2012.

CRONBACH, J. L. Coefficient alpha and the internal structure of tests. **Psychometrika**, v. 16, n. 3, p. 297-334, 1951.

DEVELLIS, R. F. Classical Test Theory. **Medical Care**, v. 44, n. 11, p. 50-59, 2006.

\_\_\_\_\_. Scale development: Theory and applications. Newbury Park, SAGE, 1991.

COELHO, E. C. Teoria da Resposta ao Item - desafios e perspectivas em exames multidisciplinares. 2014. 189f. Tese (Doutorado em Ciências) - Programa de Pós-Graduação em Métodos Numéricos em Engenharia, UFPR, Curitiba.

Organização



Promoção





GRÉGOIRE, J.; LAVEAULT, D. **Introdução às Teorias dos Testes em Ciências Humanas**. Portugal: Porto, 2002.

HALADYNA, T. M. **Developing and validating multiple-choice test items**. Lawrence Erlbaum Associates: UK, 1994.

HAYES, B. E. **Medindo a satisfação do cliente: desenvolvimento e uso de questionários**. Quality Mark: Rio de Janeiro, 1995.

HORA, H. R. M.; MONTEIRO, G. T. R.; ARICA, J. Confiabilidade em Questionários para Qualidade: Um Estudo com o Coeficiente Alfa de Cronbach. **Produto & Produção**, v. 11, n. 2, p. 85-103, 2010.

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Microdados para download**. Disponível em: <http://portal.inep.gov.br/basica-levantamentos-acessar>. Acesso em: 03 jan. 2017.

JUSTINO, G., ANDRADE, D. F. **Software para avaliação de aprendizagem utilizando a Teoria da Resposta ao Item**. In: Congresso da SBC, 27. Anais. RJ, 2007.

KELLEY, T. L. The selection of upper and lower groups for the validation of test items. **Journal of Educational Psychology**, v. 30, n. 1, p. 17-24, 1939.

LEITE, W. S. S. Sociedade moderna e tecnologias na educação: reflexões e perspectivas sobre a realidade no Brasil. **Liberato**, v. 15, n. 24, p. 105-212, 2014.

MATTHIENSEN, A. **Uso do Coeficiente Alfa de Cronbach em Avaliações**. 2011. Disponível em: <http://ainfo.cnptia.embrapa.br>. Acesso em: 14 jan. 2017.

PEREIRA, J. C. R. **Análise de dados qualitativos**. Estratégias metodológicas para as ciências da saúde, humanas e sociais. São Paulo: EDUSP/FAPESP, 2001.

R CORE TEAM. R: A Language and Environment for Statistical Computing. Vienna Austria. Disponível em: <https://www.R-project.org>. Acesso em: 05 jan. 2017.

RIZOPOULOS, D. ltm: An r package for latent variable modelling and item response theory analyses. **Journal of Statistical Software**, v. 17, n. 5, p. 1–25, 2006.

RODRIGUES, M. M. M. Proposta de Análise de Itens das provas do Saeb sob a perspectiva Pedagógica e Psicométrica. **Estudos em Avaliação Educacional**, v. 17, n. 34, 2006.

SARTES, L. M. A.; FORMIGONI, M. L. S. O. Avanços na psicometria: da Teoria Clássica dos Testes à TRI. **Psicologia: Reflex. Crítica**, v. 26, n. 2, p. 241-250, 2013.

VIEIRA, S. **Alfa de Cronbach**. Disponível em: <http://soniavieira.blogspot.com.br/2015/10/alfa-de-cronbach.html>. Acesso em: 14 jan. 2017.

Organização



Promoção





WILLSE, J. T.; SHU, Z. **CTT: Classical Test Theory Functions.** 2015. Disponível em: <https://cran.r-project.org/web/packages/CTT/CTT.pdf>. Acesso em: 05 jan. 2017.

## CLASSICAL TEST THEORY IN R ENVIROMENT AS SUPPORT IN ENGINEERING TEACHING

**Abstract:** *This paper is an exploratory research and it aims to explore some specific R packages from Classical Test Theory (CTT) like ltm and CTT with data from National Student Performance Exam for Production Engineering graduation course of 2014. Because R is free software and works with programming language it allows developing its own functions. In addition, it has a wide variety of statistical methods already implemented that are useful in test analysis. The results show that both ltm and CTT packages are stable and can be used as a solution to increase the resources used in the classroom to support in the teaching learning process. It is concluded that the reliability of the analyzed dataset is at the threshold of what is recommended in the literature and the items have little discrimination. In addition, the items that have the greatest difficulty index are those of the professional content. This paper contributes to the propagation of CTT in terms of language R allowing both educators and students to use the information here provided in their data analysis.*

**Key-words:** *Classical Test Theory, R software, ENADE, Production Engineering.*

Organização



**UDESC**  
UNIVERSIDADE  
DO ESTADO DE  
SANTA CATARINA



Promoção

