**Disciplina: Estatística Educacional TAREFA 1**

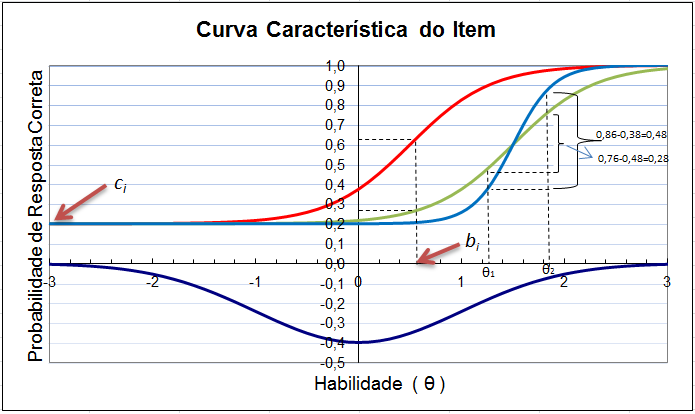
**Nas tarefas abaixo iremos trabalhar vários pontos importantes para a área de Estatísticas Educacionais. Execute-as tanto no Excel como no R (), duas ferramentas bem importantes. Em qualquer caso, quando necessário o uso de parâmetros (assim como coeficientes e constantes), atribua valores previamente (tipo a=2, b=3, D=1 etc.) no R ou coloque-as em células e atribua nomes às células ou intervalos, no Excel.**

**Google Classroom:** nomeie o arquivo de forma que ele possa ser facilmente identificado, tal como: “*Estatísticas* *Educacionais – Relatório x – Nome Sobrenome.pdf”.* Havendo mais de um arquivo, o nome deve ser o mesmo, mas com um complemento, tipo *versão 1, versão 2,...,* ou simplesmente *v1, v2,...*

**PRAZO DE ENTREGA:** *ver Google Classroom* (como todo cliente, há urgência). A Multa Contratual é de 10% por dia de atraso, se permitido.

**FORMATO DE ENTREGA:** *criar um arquivo em* ***RMARKDOWN*** *ou* ***QUARTO*** *para apresentação dos relatórios.*

1. **Gráfico de Funções**
   1. Gerar o gráfico da função *f(x)=x2-5x+6, x ε [0,4]*
   2. Gerar o gráfico da função *f(x)=densidade da N(0,1) e da Função de Distribuição da N(0,1).*
   3. Gerar o gráfico da função *f(x)=1/(1+exp(-Da(x-b)), com a=1,5, b=1 para D=1 e D=1,7*
   4. Compare a função de distribuição obtida em (1.2) com a função obtida em (1.3) com D=1,7
   5. Gerar o gráfico da função *f(x)=c+(1-c)/(1+exp(-Da(x-b)), com a=1,5, b=1, c=0,2 e D=1,7*
   6. Gerar o gráfico da função *f(x)=densidade da N(0,1) refletida (-f(x)),* junto com 3 outras funções logísticas de parâmetros *ζ =(a,b,c)* dados por (1 ,.5, 0.2), (1, 1.5, 0.2) e (2, 1.5, 0.2) [ver exemplo abaixo]



1. **Geração de Dados Simulados**
   1. Gerar n=1000 valores de uma v.a. U(0,1) e plotar o Histograma.
   2. Gerar n=1000 valores de uma v.a. , com p=0,3. Verifique se os valores teóricos (média e variância) estão próximos das Estatísticas. [Lembrar que E(X)=p e Var(X)=p(1-p)]
   3. Refaça o item 2.2 considerando a Bin(n, p), com n=10 e p=0,5.
   4. Refaça o item 2.2 considerando a N(0,1).

**Dica**: Para gerar observações de uma U(0,1) no Excel use o comando =***ALEATORIO()*** na célula desejada, e para gerar no R use ***runif(1)***. Para N(0,1) use =***INV.NORMP.N(ALEATÓRIO())*** no Excel e ***rnorm(1)*** no R. Para gerar Bernoulli(p) no Excel use ***=SE(ALEATORIO()<p,1,0)*** ou ***=1\*(ALEATORIO()<p)*** e ***rbinom(1,1,p)*** no R. Também pode-se usar o ***Suplementos de Análise*** do Excel.

1. **Simulando dados de avaliação**
   1. Gerar n=1000 valores de uma N(0,1) representando as habilidades dos indivíduos (x) na coluna A do Excel
   2. Usando a expressão e valores em (1.5), calcular as probabilidades de os *n* indivíduos acertarem ao item (Coluna B).
   3. Com as probabilidades em (3.2), gerar valores de *n* Bernoullis (Coluna C)
   4. Acompanhe o arquivo **TRI-Respostas.xlsx** e **GeraDados.xlsx** para geração por fórmulas ou VBA.
2. **Otimização: obtenção do máximo de uma função**
   1. Use o **Suplemento** **Solver** do Excel para obter o valor que maximiza a densidade de uma N(0,1)
   2. Usando a função de probabilidade *P(U=x)=px(1-p)1-x* e uma amostra (x1,...,x10) = (0,1,0,0,1,0,0,0,1,0) construa a função de verossimilhança (produto das probabilidades). Use o Suplemento Solver para obter o valor de *p* que maximiza a verossimilhança (isto é, o Estimador de Máxima Verossimilhança).
3. **Gerenciamento de Bases de Dados no EXCEL**
   1. Baixe os arquivos ***ENEM2017MT.CSV*** e ***ENEM2017MT\_infoitem.CSV*** em ***tri.helitontavares.com***, pasta **TAREFAS/TAREFA1/ENEM2017**. Eles contêm as respostas e gabaritos do **ENEM-2017**, **Caderno AZUL**, com respostas a 45 itens de **MATEMÁTICA (MT)**, bem como informações dos itens. Os demais arquivos, correspondentes às áreas de **CN**, **CH** e **LC**, estão no mesmo site.
   2. Importe o arquivo de **respostas** e coloque uma primeira planilha do Excel, separando em colunas.
   3. Importe o arquivo de **gabarito** e coloque uma segunda planilha, separando em colunas.
   4. Crie uma **terceira** planilha para colocar os **dados corrigidos** (1: acerto, 0: qualquer outra), usando a macro VBA a seguir:

|  |
| --- |
| Sub corrige()  nindiv = 82233 'Ajuste o número. Use 10 para testes  itens = 45  'Lembre-se que Sheets=Planilha (matriz) e que Cells=Célula (elemento) de uma matriz  For n = 1 To nindiv  For i = 1 To itens  If (Sheets(1).Cells(n + 1, i + 1) = Sheets(2).Cells(i + 1, 2)) Then Sheets(3).Cells(n + 1, i + 1) = 1 Else Sheets(3).Cells(n + 1, i + 1) = 0  Next  Next  End Sub |

PS1. Lembre-se que pode-se abrir o editor VBA através: (i) menu Exibir > Macros, (ii) menu Desenvolvedor > Macros, (iii)

PS2. Salve o arquivo com macro com extensão **\*.xlsm** (Pasta de Trabalho Habilitada para Macro), ou perderá a macro

* 1. Copie os IDs dos indivíduos e os Códigos dos itens para a planilha corrigida. Essa parte também poderia ser incluída na macro.
  2. Salve a planilha em arquivo TEXTO no formato exemplificado abaixo (sem a linha de códigos). Este formato poderá ser lido em alguns programas específicos de Avaliação Educacional. No exemplo, os 12 primeiros caracteres são destinados ao identificador (ID) dos indivíduos, um espaço separa os IDs dos dados 01.

Dica: No Excel, use a opção ***Salvar como*** e salve os dados no formato ***Texto formatado (separado por espaços)***. Ajuste a largura das colunas para obter o resultado desejado. Salve com o nome ***ENEM2017\_01.TXT***

|  |
| --- |
| 170000007404 000001100000101000100111000010100000000100000  170005101416 000010000000000010010010001000100100000100010  170000098218 101011000011101011011010001111001000011100100  170005205041 100011011111001011010111111111000011011101111  170005203958 100111000100100000001111000010000010100110010  170000162158 111111111011100111011110011111101110101100110  170000147521 101011110111101111111110011000101111101111111  170000182413 100011000010000000000011110010100001001000000  170001874145 100011000001101001110011001010101010011100110  170003535712 111001111011000111111111001010100000000100000 |

1. **Gerenciamento de Bases de Dados no R**
   1. Refaça o procedimento (leitura de dados e gabaritos, correção para 0 e 1) usando o R. Abaixo alguns códigos:

|  |
| --- |
| #setwd("C:\\Users\\...\\TAREFAS"); getwd()  # LEITURA DO ARQUIVO DE INFORAÇÃO DOS ITENS  ENEM2017MT\_infoitem = read.csv("ENEM2017MT\_infoitem.csv", sep=";")[,]  # LEITURA DO ARQUIVO COMO STRING DE RESPOSTA  ENEM2017MT\_1 = read.csv("ENEM2017MT.csv", sep=";")  # LEITURA EM COLUNA SEPARADAS  ENEM2017MT = read.fwf('ENEM2017MT.csv', widths=c(12,-1,rep(1,45)), skip=1,header=FALSE)  #ATRIBUIÇÃO DOS NOMES DAS VARIÁVEIS  colnames(ENEM2017MT) = c('ID', ENEM2017MT\_infoitem[,1]) |

|  |
| --- |
| nindiv=nrow(ENEM2017CH); nitens=ncol(ENEM2017CH)-1  CODitens=ENEM2017CH\_infoitem[,1]  GAB=ENEM2017CH\_infoitem[,2]  ENEM2017CH\_01 = data.frame(matrix(NA,nindiv,nitens)); dim(ENEM2017CH\_01)  colnames(ENEM2017CH\_01) = CODitens    for (j in 1:10) # nindiv  for (i in 1:nitens)  ENEM2017CH\_01[j,i]=1\*(ENEM2017CH[j,i+1]==GAB[i])  # for (j in 1:nindiv) ENEM2017CH\_01[j,] = 1\*(ENEM2017CH[j,-1]==GAB) # forma alternativa |

* 1. Dicotomizar (corrigir) as respostas para 01.
  2. Obtenha os **Escores** (Número de Acertos) e plote o **Histograma** e comente.
  3. Para cada item, obtenha a **Proporção de Acertos (PA)** (pode ser com o ***colMeans***). Plote o histograma e comente.

1. **No link *tri.helitontavares.com*, pasta TAREFAS/RAI, baixe o arquivo RAI0008 - ENEM-2017 - MT.PDF** (feito em RMarkdown)
   1. Monte um código no R para gerar um gráfico da Seção 3 em que x=Escore e y=PA para os 5 primeiros itens.
2. Baixe os **arquivos** ***ENEM2017.CSV*** e ***ITENS\_PROVA\_2017.CSV***, que são similares ao microdado do **ENEM**, mas apenas com o Caderno **AZUL** de dados.Neste, constam as 4 áreas com as respostas e os respectivos gabaritos em cada linha.
   1. Monte o código para fazer a leitura no R, separação e correção para a área de LC.
   2. Plote os histogramas para as notas nas 4 áreas e em Redação.

**Bom trabalho!**