

Estatística e Probabilidade - Projeto 02 - 2022/01

Prof. Hugo Carvalho

21/07/2022

– INSTRUÇÕES – LEIAM ATENTAMENTE! –

- A data limite de entrega da avaliação é sábado 06/08/2022 às 23h59. Avaliações entregues após esse prazo serão desconsideradas.
- A entrega deve ser feita exclusivamente através do Google Classroom, da seguinte forma:
 - Dentro da turma no Google Classroom, clique em “Atividades”
 - No tema “Projetos”, clique em “Projeto 02” e em seguida em “Ver atividade”
 - No canto superior direito, clique em “+ Adicionar ou criar” e em seguida clique em “Documentos”. Você também pode anexar um notebook do Google Colaboratory, clicando em “Google Drive” e selecionando o notebook correspondente. As instruções a seguir são majoritariamente para relatórios feitos no Google Docs.
 - Você deverá fazer o trabalho no documento que irá abrir em uma nova guia. O título do documento será automaticamente “SEU NOME - Projeto 02”. Por favor não o altere!
 - Você pode fechar essa guia para retomar o trabalho depois. O documento é salvo automaticamente pelo Google Docs, caso haja conexão com a Internet. Fique atento/a à informação de “Salvando...” ou “Alterações salvas no Drive” no topo da página, logo após fazer qualquer alteração no texto.
 - Caso você queira fechar o trabalho e retomá-lo depois, para voltar do ponto que parou vá em “Atividades > Projetos > Projeto 02 > Ver atividade”. No canto superior direito da tela, acima do botão “+ Adicionar ou criar”, você verá um arquivo chamado “SEU NOME - Projeto 02”. Clique nele, e continue do ponto em que parou.
 - Para entregar o projeto, abra o documento que o contém (conforme procedimento explicado no ponto acima), e no canto superior direito clique em “ENTREGAR” (o texto estará em uma caixa cinza, mas ele é clicável).
 - Na tela a seguir clique em “Entregar” para confirmar a entrega do trabalho.
 - Se você se arrependeu e quiser fazer outra entrega, basta abrir a atividade, no canto superior direito clique em “Cancelar envio” e confirme clicando novamente em “Cancelar envio” na janela que irá abrir. Você terá acesso ao seu documento de onde parou e poderá fazer a entrega novamente até o prazo determinado.
- O corpo do texto deve estar escrito em tipografia Georgia ou EB Garamond, de tamanho 11, com espaçamento de 1,5 e texto justificado.
- Trechos de código devem estar escritos em tipografia Courier New, de tamanho 11, com espaçamento simples e texto alinhado à esquerda. Não é necessário estar com coloração como é usual nos editores de texto.
- Caso seja necessário inserir equações, utilize o módulo de equações do Google Docs (disponível em “Inserir > Equação”).
- A troca de conhecimento na realização da avaliação é permitida e encorajada: ciência se faz com colaboração, e devemos seguir esse espírito aqui. Porém, cada aluno deverá ter a sua própria resolução, e cópias ou outras ilegalidades serão severamente punidas com a anulação da avaliação para o(s) aluno(s) suspeito(s).
- Todos os passos de sua resolução devem ser devidamente justificados.
- Ao entregar essa avaliação, você afirma ter lido e estar de acordo com essas regras, comprometendo-se a cumpri-las.

Questão 1: Disserte sobre a *distribuição de Zipf* (ou *lei de Zipf*, como também é chamada). Seu texto deve falar sobre: motivação e surgimento da distribuição, algumas de suas diversas aplicações, trazer discussões/interpretações de seus sumários numéricos (média, variância, etc.) e tratar de outros aspectos probabilísticos da distribuição que você julgue importante. Finalmente, diga como que a distribuição de Zipf se aplica em compressão de arquivos. Faça um texto o mais detalhado possível, evitando frases “misteriosas”; coloque referências bibliográficas para afirmações que você não irá se aprofundar.

Questão 2: O objetivo desta questão é ilustrar que certas probabilidades podem ser calculadas malandramente de modo levemente indireto.

- Seja Z uma variável aleatória com distribuição normal padrão, ou seja, normal de média zero e variância unitária. Argumente que $\mathbb{P}(Z > 20)$ não é uma quantidade nula mas que é impossível obtê-la “ingenuamente”, ou seja, através de algum dos seguintes métodos: consultando uma tabela, utilizando as funções elementares de algum *software* ou pacote (p. ex., `import scipy.stats` e em seguida `1 - scipy.stats.norm(0, 1).cdf(20)` no Python), ou através de simulação estocástica (p. ex., calculando a proporção de observações acima de 20 em uma grande quantidade de simulações de tal variável aleatória).
- Escreva a integral que calcula a quantidade $\mathbb{P}(Z > 20)$, faça uma mudança de variáveis esperta e conclua que tal probabilidade pode ser expressada como $\mathbb{E}[g(Y)]$, onde Y é uma variável aleatória com distribuição uniforme no intervalo $[0, 1/20]$ e a função g é dada por

$$g(y) = \frac{1}{20y^2\sqrt{2\pi}} e^{-1/2y^2}.$$

- Usando o resultado do item b) encontre uma aproximação para $\mathbb{P}(Z > 20)$ através de simulações da variável aleatória Y .
- Ao se repetir o processo do item c) diversas vezes, você obterá várias estimativas para $\mathbb{P}(Z > 20)$. Estime a incerteza em tais estimativas, justificando o seu procedimento. Discuta o que acontece com tal incerteza à medida que a quantidade de simulações de Y aumenta.

Questão 3: O objetivo desta questão é estudar dados provenientes de um estudo médico com olhar probabilístico. Ao se testar um novo tratamento para determinada doença, 10 pacientes enfermos foram sujeitos ao tratamento e observou-se que 7 ficaram curados e 3 não. Com base nisso, deseja-se averiguar a *probabilidade de sucesso* do tratamento, ou seja, a chance de um paciente ficar curado quando submetido a tal tratamento.

- Argumente que se a probabilidade de sucesso do tratamento é dada por s , então a probabilidade de se observar exatamente k curas no tratamento de 10 enfermos pode ser dada pela quantidade abaixo, justificando as devidas hipóteses para tal:

$$\binom{10}{k} s^k (1-s)^{10-k}.$$

- Após realizar estudos teóricos, um cientista afirma categoricamente que a probabilidade de sucesso deste tratamento é de 35%. Para confrontar tal afirmativa com os dados experimentais use o resultado do item a) para calcular numericamente a probabilidade de se observar 7 ou mais sucessos no tratamento dos 10 pacientes, assumindo que a probabilidade de sucesso do tratamento é de fato de 35%. Você acha que tal valor é compatível com a afirmação do cientista? Justifique sua resposta.
- Para fazer um tratamento *via* Inferência Bayesiana de tal problema, interprete a probabilidade de sucesso do tratamento também como uma variável aleatória. Assuma, de modo conservador, que previamente à realização de algum experimento a probabilidade de sucesso do tratamento assume os valores abaixo com igual probabilidade:

$$\{0, 0,01, 0,02, \dots, 0,97, 0,98, 0,99, 1\}.$$

Denote por \mathcal{D} os dados observados, e usando o Teorema de Bayes mostre que a nova distribuição de probabilidade da probabilidade de sucesso, após realização do experimento, é atualizada para

$$\mathbb{P}(S = s|\mathcal{D}) = \frac{\mathbb{P}(\mathcal{D}|S = s)\mathbb{P}(S = s)}{\mathbb{P}(\mathcal{D})} = \frac{[s^7(1-s)^3]}{\sum_s [s^7(1-s)^3]},$$

para $s = 0, 0,01, 0,02, \dots, 0,99, 1$.

- Com base nesta informação, calcule a probabilidade de que a probabilidade de sucesso do tratamento seja maior que 35%. O que você pode dizer ao comparar tal valor com a afirmativa do cientista do item b)?