

Lab06 - Estatística Inferencial II

Aula para sanar dúvidas: dia 20/03/2012 das 18 às 20 horas no LCC2.

Limite de entrega sem penalização: 22/03/2012 às 23:59.

Liberação do gabarito: 26/03/2012.

Problema

O jovem Sidnelson, um ativo habitante do mundo virtual (a.k.a. *nerd*), está se mudando para a Paraíba em breve. Como usuário assíduo de jogos em rede, uma das suas principais preocupações na mudança é com a velocidade da conexão de Internet que ele terá na sua nova casa. Ele descobriu dois provedores na Paraíba que fornecem conexão com a velocidade que ele deseja (50Mbps): o provedor “GZT” e o provedor “Ola” Internet, ambos pelo mesmo preço.

Por sorte, ele conheceu pelas redes sociais dois vizinhos da sua nova moradia, dos quais um utiliza o provedor “GZT” e o outro o provedor “Ola”, ambos com o mesmo plano de 50Mbps. Não satisfeito com os elogios feitos pelos vizinhos sobre as conexões, ele queria verificar na prática qual dos dois provedores fornecem a melhor conexão para que ele escolhesse qual contratar. Além disso, ele queria saber se os provedores realmente cumprem algumas coisas que prometem dos seus serviços. Assim, Sidnelson pediu para que seus dois vizinhos instalassem e executassem um medidor de velocidade, que testa a velocidade da conexão com a Internet de hora em hora, durante um mês.

Apesar de *nerd*, o jovem Sidnelson não está cursando a disciplina de *Analytics* e está pedindo aos colegas que analisem os dados coletados e que o ajudem a tomar a decisão de que provedor escolher, além de verificar algumas promessas feitas pelos provedores.

Instruções

- Para cada questão, identifique quais os testes mais adequados para tirar suas conclusões e argumente como você fez suas escolhas. Identifique também quais as estatísticas e os *plots* mais adequados para representar os dados e ajudar nas discussões;
- As pré-condições para utilizar cada teste devem ser verificadas e discutidas;
- Você deve escrever um relatório em **um** arquivo *.pdf*, com as análises feitas em todas as questões, apresentando com detalhes como você chegou a cada conclusão. Além disso, o relatório **deve** apresentar figuras e resultados numéricos de testes realizados para apoiar suas análises;
- Para cada questão, você **deve** escrever um script R que execute todos os testes usados para resolver a questão e que gere como saída arquivos texto contendo os resultados dos testes usados na análise em formato de tabela, além de gerar também as figuras usadas nas discussões;

- Os dados do arquivo de entrada, que podem ser encontrados no site da disciplina em <https://sites.google.com/a/computacao.ufcg.edu.br/analytics/curriculo/fundamentos-de-analytics/labs/velocidade-isp.txt?attredirects=0&d=1>, possuem as seguintes colunas:
 - **ano** [*integer*]: ano no qual a medição de velocidade foi realizada;
 - **mes** [*character*]: mês no qual a medição de velocidade foi realizada;
 - **dia** [*integer*]: dia do mês no qual a medição de velocidade foi realizada;
 - **hora** [*integer*]: hora do dia na qual a medição de velocidade foi realizada;
 - **velocidade** [*double*]: velocidade da conexão (em Mbps) medida na hora indicada;
 - **provedor** [*character*]: nome do provedor de internet responsável pela conexão medida;
 - **estado** [*character*]: nome do estado onde foi realizada a medição.
- Assuma que as medições que possuem valores iguais para o ano, dia e hora foram feitas exatamente no mesmo instante.

Com base nos dados de testes de velocidade e seguindo as instruções acima, resolva as questões a seguir.

Questões

1. Disseram a Sidnelson que na prática os dois provedores da Paraíba possuem, em média, a mesma velocidade de conexão para o plano de 50Mbps. Verifique se o que dizem realmente é verdade, usando as medições feitas de hora em hora para os provedores “GZT” e “Ola” da Paraíba (estado == “PB”), que foram realizadas no mês de Dezembro de 2011.

Discuta se existe diferença estatística entre as médias de velocidade de conexão dos dois provedores. Caso exista diferença, qual provedor você indicaria a Sidnelson? Faça uma análise para diferentes níveis de significância.

2. Lendo os contratos dos provedores “GZT” e “Ola” da Paraíba, Sidnelson viu que ambos prometem que a média mensal da conexão será sempre maior ou igual a 50Mbps. Verifique se esta promessa se confirma, usando as medições feitas de hora em hora para os dois provedores da Paraíba (estado == “PB”).

3. Sidnelson também viu que no contrato da “GZT” eles dizem que a qualidade de serviço oferecida em todas regiões do país é a mesma para planos semelhantes, incluindo a velocidade de conexão. Com seus contatos nas redes sociais, Sidnelson pediu pra dois amigos, um do Rio de Janeiro e outro de São Paulo, fazerem a mesma medição de hora em hora na velocidade da conexão de seus provedores, ambos “GZT”. Porém, as medições foram feitas em meses diferentes, como consta nos dados.

Verifique se é possível afirmar estatisticamente que a média da velocidade de conexão da “GZT” nos três estados medidos (PB, RJ e SP) realmente não possuem diferença estatística, como é dito no contrato da empresa. As diferenças nos valores de velocidade de conexão são variações normais de fatores internos, ou realmente há diferença entre a velocidade de conexão da “GZT” entre os estados?

Caso seja identificada diferença de velocidade entre os estados, analise estatisticamente e identifique qual o estado possui a melhor conexão, discutindo como você chegou à sua conclusão.