

Universidade Federal do Paraná - Departamento de Estatística

Disciplina: CE314 – Modelos Lineares Generalizados

Prof. Cesar Augusto Taconeli

Orientação - Trabalhos

Os alunos deverão constituir **grupos com três componentes**. Serão dois trabalhos, um sobre regressão para dados binários e outro para dados de contagens, **que deverão ser apresentados na forma de seminários**.

Cronograma

10/10	Data limite para encaminhar por e-mail (cetaconeli@gmail.com) a composição dos grupos ou as informações dos nomes sem grupos.
13/10	Definição dos temas atribuídos a cada grupo, a ser divulgada no ambiente da disciplina no moodle
18/11	Seleção dos grupos debatedores para as apresentações
21/11	Data limite para os grupos encaminharem os slides, em formato PDF, para o email cetaconeli@gmail.com
25/11, 02/12 e 05/12	Apresentação dos seminários

Trabalho 1– Regressão para dados binários

- Os dados deverão ser extraídos da página do Departamento de Informática do Sistema Único de Saúde (DATASUS). O link segue abaixo:

<https://datasus.saude.gov.br/informacoes-de-saude-tabnet/>

- Para extração da variável resposta, você deverá clicar no link “**Demográficas e sócio-econômicas**”. Clique em “Trabalho e renda- Censos 1991, 2000 e 2010” ou “Educação- Censos 1991, 2000 e 2010”, conforme a variável resposta utilizada de acordo com alocação aos grupos (ver abaixo).
- Selecione a unidade da federação no Mapa conforme alocação aos grupos (ver abaixo). Selecione “Município” na linha, a variável de interesse na coluna (ou conteúdo) e o ano de 2010.
- Selecione a opção “Exibir linhas zeradas” e clique em “Mostra”.

- As unidades da federação serão atribuídas aos grupos, da seguinte forma:
 - **Grupo 1:** Rio Grande do Norte- Renda média domiciliar per-capita (criar a variável resposta binária: abaixo do primeiro quartil vs acima do terceiro quartil e modelar a chance de “alta renda”);
 - **Grupo 2:** Pará- Índice de Gini (criar a variável resposta binária: abaixo do primeiro quartil vs acima do terceiro quartil e modelar a chance de “elevada renda”);
 - **Grupo 3:** Santa Catarina- Razão de renda (criar a variável resposta binária: abaixo do primeiro quartil vs acima do terceiro quartil e modelar a chance de “elevada razão de renda”);
 - **Grupo 4:** Rio Grande do Sul- Proporção de pessoas com baixa renda (criar a variável resposta binária: abaixo do primeiro quartil vs acima do terceiro quartil e modelar a chance de “elevada proporção com baixa renda”);
 - **Grupo 5:** Goiás- Proporção de crianças em situação domiciliar de baixa renda (criar a variável resposta binária: abaixo do primeiro quartil vs acima do terceiro quartil e modelar a chance de “elevada proporção crianças em situação com baixa renda”);
 - **Grupo 6:** Rio de Janeiro – Taxa de desemprego (criar a variável resposta binária: abaixo do primeiro quartil vs acima do terceiro quartil e modelar a chance de “elevada taxa de desemprego”);
 - **Grupo 7:** Piauí- Taxa de trabalho infantil (criar a variável resposta binária: abaixo do primeiro quartil vs acima do terceiro quartil e modelar a chance de “elevada taxa de trabalho infantil”);
 - **Grupo 8:** São Paulo- Taxa de analfabetismo (criar a variável resposta binária: abaixo do primeiro quartil vs acima do terceiro quartil e modelar a chance de “elevada taxa de analfabetismo”);
 - **Grupo 9:** Mato Grosso- Escolaridade da população de 15 anos ou mais, sem instrução/1º ciclo fundamental incompleto (criar a variável resposta binária: abaixo do primeiro quartil vs acima do terceiro quartil e modelar a chance de “elevada taxa de baixa istrução”).
- Como variáveis explicativas (preditores), deverão ser selecionadas variáveis demográficas e socioeconômicas, também disponíveis no link do DATASUS. Verifiquem as variáveis disponíveis e escolham, aproximadamente, 4 e 8 delas para a análise. Caso essas variáveis não estejam disponíveis para o ano de 2010, escolha o ano anterior mais recente disponível. **Nota:** Nem todas as variáveis precisarão ser usadas no modelo final.
- Cada grupo deverá desenvolver o trabalho baseado em análise de regressão logística. A apresentação deverá contemplar, **de maneira concisa**:
 - Definição do problema e descrição resumida dos dados e das variáveis consideradas;
 - Análise descritiva/exploratória;
 - Especificação do modelo ajustado;

- Resumo dos resultados do modelo ajustado;
 - Diagnóstico do ajuste (análise de resíduos, observações atípicas, multicolinearidade e teste da qualidade do ajuste);
 - Interpretação/discussão dos efeitos das preditores na resposta;
 - Conclusão do trabalho, apresentando o fechamento da análise, as limitações encontradas e possíveis investigações futuras.
- Se adequado, algumas das variáveis explicativas podem ser categorizadas para a análise (usando pontos de corte, ou dividindo conforme tercis ou quartis, por exemplo), ou transformadas de outra forma (por exemplo, usando transformação logarítmica ou Box-Cox caso a distribuição seja esparsa e assimétrica).
 - Como as populações dos municípios têm tamanhos diferentes, o número de habitantes (ou seu logaritmo) deverá considerado no ajuste do modelo. Recomenda-se que as demais variáveis sejam inseridas na forma de taxas, para evitar multicolinearidade.

Trabalho 2– Regressão para dados de contagens

- Os dados deverão ser extraídos de da página do Departamento de Informática do Sistema Único de Saúde (DATASUS). O link segue abaixo:

<https://datasus.saude.gov.br/informacoes-de-saude-tabnet/>

- Para extração da variável resposta, você deverá clicar no link “**Estatísticas vitais**”. Clique em “Mortalidade – desde 1996 pela CID10” e depois em “Óbitos por causas externas”.
- Selecione a unidade da federação no Mapa. Selecione “Município” na linha, “Não ativa” na coluna e “Óbitos por ocorrência” no conteúdo, e o ano de 2023;
- Na aba “Grande grupo CID10, escolha uma das categorias (Acidentes de transporte, outras causas externas...).
- Selecione a opção “Exibir linhas zeradas” e clique em “Mostra”;
- As unidades da federação serão atribuídas aos grupos, da seguinte forma:
 - Grupo 1: Rio Grande do Norte;
 - Grupo 2: Pará;
 - Grupo 3: Santa Catarina;
 - Grupo 4: Rio Grande do Sul;
 - Grupo 5: Goiás;
 - Grupo 6: Rio de Janeiro;
 - Grupo 7: Piauí;
 - Grupo 8: São Paulo;
 - Grupo 9: Mato Grosso.
- Como variáveis explicativas (preditores), deverão ser selecionadas variáveis demográficas e socioeconômicas, também disponíveis no link do DATASUS. Verifiquem as variáveis disponíveis e escolham, aproximadamente, 4 e 8 delas para a análise. Caso essas variáveis não estejam disponíveis para o ano de 2023, escolha o ano mais recente disponível. **Nota:** Nem todas as variáveis precisarão ser usadas no modelo final.
- Cada grupo deverá desenvolver o trabalho baseado no ajuste de MLGs para contagens. A apresentação deverá contemplar, **de maneira concisa**:
 - Definição do problema e descrição resumida dos dados e das variáveis consideradas;
 - Análise descritiva/exploratória;

- Especificação do modelo ajustado;
 - Resumo dos resultados do modelo ajustado;
 - Diagnóstico do ajuste (análise de resíduos, observações atípicas, multicolinearidade e teste da qualidade do ajuste);
 - Interpretação/discussão dos efeitos das preditores na resposta;
 - Conclusão do trabalho, apresentando o fechamento da análise, as limitações encontradas e possíveis investigações futuras.
- Se adequado, algumas das variáveis explicativas podem ser categorizadas para a análise (usando pontos de corte, ou dividindo conforme tercis ou quartis, por exemplo), ou transformadas de outra forma (por exemplo, usando transformação logarítmica ou Box-Cox caso a distribuição seja esparsa e assimétrica).
- Como as populações dos municípios têm tamanhos diferentes, o número de habitantes (ou seu logaritmo) deverá considerado no ajuste do modelo. Recomenda-se que as demais variáveis sejam inseridas na forma de taxas, para evitar multicolinearidade.
- A estrutura da apresentação deverá ser semelhante à sugerida para o caso da regressão para dados binários. Para a análise dos dados de contagens, no entanto, recomenda-se particular cuidado na avaliação de possível super-dispersão e o uso de métodos adequados, se for o caso.

Apresentação

- Metade dos grupos apresentará, na forma de seminário, o trabalho desenvolvido para a análise dos dados binários; a outra metade apresentará a análise de dados de contagens. **Todos deverão, no entanto, encaminhar por e-mail os slides referentes às duas apresentações.**
- As apresentações ocorrerão nas datas especificadas no cronograma;
- Cada grupo terá de **10 a 15 minutos** para sua apresentação (apresentações com duração fora desse intervalo terão suas notas descontadas);
- Haverá **5 a 10 minutos** adicionais para questionamentos **por parte de um grupo debatedor**, que será selecionado previamente;

Algumas sugestões:

- Utilizar 15 a 20 slides, evitando exceder o tempo definido para a apresentação e permitindo apresentar o artigo num ritmo adequado;
- **Inserir pouco texto nos slides.** Utilizar, sempre que possível, tópicos curtos. Complementar as informações projetadas oralmente;
- **Evitar a simples leitura dos slides;**
- Ao término da apresentação, responder com boa vontade às questões dos colegas.