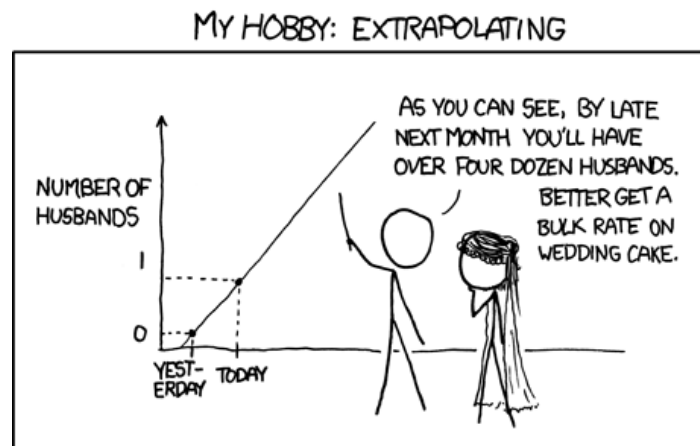


Projeto Final De Disciplina

Universidade Federal da Paraíba
Ciência de Dados para Negócios
Disciplina: Econometria I
Docente: Antonio Vinicius Barbosa

Data de entrega: 28 de outubro de 2024



Este projeto está dividido em duas partes. Leia atentamente o que se pede e forneça respostas diretas e precisas. Lembre-se de especificar corretamente o caminho onde os dados estão armazenados no seu computador.

1 Discriminação no mercado de trabalho

Um ramo da Economia concentra-se no estudo dos ganhos salariais no mercado de trabalho. O que faz um indivíduo ganhar mais do que outro? Citando Adam Smith em A Riqueza das Nações (livro I, Cap. 10): “Os salários aumentam com a dificuldade e o custo de aprender uma tarefa”. Isso estabelece uma ligação teórica entre a educação e os salários: a educação aumenta a produtividade do trabalhador e, conseqüentemente, seu salário.

Uma ferramenta para estudar empiricamente a relação entre a educação e os salários são as chamadas equações de renda ou equações de Mincer¹. A equação 1 é a especificação mais básica de uma equação *minceriana*:

$$\log(\text{salario}_i) = \beta_0 + \beta_1 \text{educ}_i + \beta_2 \text{homem}_i + e_i \quad (1)$$

onde $\log(\text{salario}_i)$ representa o logaritmo dos rendimentos por hora ou taxa salários por hora do trabalhador i , educ_i denota seus anos de educação formal, homem_i é uma variável *dummy* que indica 1 se for um homem e 0 caso contrário, e e_i é um termo aleatório que supomos que segue as suposições clássicas.

- Baixe e carregue os dados `salario_mercado.csv`.

Agora, responda as seguintes perguntas:

1. Rodar o modelo descrito na equação 1.
2. Calcular o efeito da variável *educ* o salário e interpretar seu significado. Qual sinal esperaríamos obter desta relação?
3. Calcule as médias salariais para homens e mulheres.
4. Vamos nos concentrar nas disparidades salariais entre homens e mulheres que compartilham as mesmas características observáveis. Com base na equação 1, obter a expressão que indica a diferença média na taxa horária de salário de um homem em comparação com uma mulher com características idênticas. Propor uma hipótese nula e alternativa para avaliar a ausência de disparidade salarial entre homens e mulheres. Em seguida, suponha que $\beta_2 > 0$. Há, nos dados, evidência suficiente para afirmar que existe discriminação contra as mulheres no mercado de trabalho?

2 Taxa de glicose e incidência de diabetes

Nesta parte, será analisado um conjunto de dados com informações de saúde de uma determinada população, como parte de um estudo para prever diabetes em mulheres adultas (com 21 anos ou mais).

As seguintes variáveis estão incluídas nos banco `diabetes_dados.csv`:

- `n_gravidez`: Número de vezes que a pessoa ficou grávida
- `glicose`: Concentração de glicose plasmática 2 horas após um teste oral de tolerância à glicose
- `pressao_arterial`: Pressão arterial diastólica (mm Hg)
- `insulina`: Nível de insulina sérica após 2 horas (mu U/ml)
- `imc`: Índice de massa corporal (peso em kg/(altura em m)²)
- `idade`: Idade (anos)
- `diabetes`: Variável de dummy (0 = sem diabetes ou 1 = com diabetes)

¹Mincer, J. (1974). Schooling, experience, and earnings. National Bureau of Economic Research. Columbia University Press. New York.

Agora, siga o roteiro abaixo:

1. Apresente uma tabela descritiva de dados contendo média e desvio padrão para cada variável
2. Faça um gráfico relacionando idade e o nível de glicose. Em seguida, relacione pressão arterial e o IMC. Quais conclusões podemos tirar dessas relações?
3. Agora, estime um modelo de regressão para as duas relações anteriores. Discuta detalhadamente os resultados dos estimadores e as conclusões obtidas. Para os dois casos, construa o intervalo de confiança a 95% utilizando a função `confint()` e discuta os resultados.
4. Visualize o modelo estimado através da adição da reta de ajuste ao gráfico de dispersão plotado anteriormente.
5. Temos uma variável chamada `n_gravidez`, que fornece o número de gestações anteriores. Crie uma variável dummy chamada `d_gravidez`: qualquer mulher com 1 ou mais gestações será codificada como 1; e qualquer mulher sem histórico de gravidez será codificada como 0.
6. Agora rode o modelo com o nível de glicose como variável dependente e idade e a dummy de gravidez como variáveis independentes. Como podemos interpretar os coeficientes e quais conclusões podemos tirar do resultado da estimação?

3 Avaliação

Os alunos deverão entregar um arquivo PDF contendo as respostas, assim como os códigos utilizados. A seguir são apresentados os critérios de avaliação do projeto:

1. Aspectos gerais

Cada projeto será avaliado de acordo com alguns critérios, tais como:

- Entrega no prazo estabelecido
- Análise relevante
- Redação e clareza
- Impressões gerais

2. Visualização

Serão avaliados os seguintes critérios:

- Uso de gráficos apropriados
- Aparência geral
- Relevância da informação

3. Código

As duplas devem reportar todos os códigos utilizados no projeto. Qualquer análise/gráfico apresentado sem o devido código em R ou Python será considerado incompleto. As seguintes regras gerais serão observadas:

- Clareza e facilidade de leitura
- O código é comentado
- Disponibilização do código-fonte

4 Prazo de entrega

A entrega do projeto deverá ser feito em **duplas** no dia **28/10/2024**, impreterivelmente. Deseja-se que o grupo envie uma pasta compactada (contendo todos os arquivos - PDF e códigos) através da página do Sigaa até esta data limite.