

Trabalho Final – Estatística – 2024

Instruções:

O trabalho é **opcional**. A nota do trabalho substituirá a menor nota, entre a A1 e a A2.

O trabalho pode ser feito individualmente ou em duplas. Caso feito em duplas, é esperado que ambos os integrantes contribuam igualmente para o resultado.

A data de entrega é **sexta-feira, dia 22/11**. Uma versão física impressa e colorida do relatório deve ser entregue durante a aula do dia 22/11, e uma versão virtual do mesmo relatório deve ser enviada por email no mesmo dia. Por favor, envie o trabalho para bruno.barsanetti@fgv.br. O título do email deve ser “[Estatística] Trabalho Final”.

Os dados para o trabalho estão na planilha em anexo. Os dados são dos censos demográficos de 2000 e 2010. Um breve dicionário abaixo:

cod_micro: o código usado pelo IBGE para indicar a microrregião.

nome_micro: o nome da microrregião, em geral o nome do principal município.

empregoYYYY: o total de pessoas de mais de 10 anos de idade que trabalharam na semana anterior, medido no censo do ano YYYY (2000 ou 2010). Essa variável mede o número de empregos em uma microrregião.

rendaYYYY: a renda mensal média das pessoas empregadas no censo do ano YYYY. Essa variável mede principalmente o rendimento do trabalho, que tendem a refletir diferenças de salários.

Microrregiões são grupamentos regionais de municípios economicamente interligados. Em economia, frequentemente as microrregiões são usadas como as unidades onde mercados de trabalho locais operam. Isto é, costuma-se presumir que todas as firmas e empregados em uma microrregião estão operando no mesmo mercado de trabalho. Use as microrregiões como a unidade de análise.

O trabalho pode ser feito em qualquer software estatístico, inclusive o Excel (que eu recomendo), e não há entrega de códigos. O relatório deve ser claro e bem-escrito, a escrita sucinta, e os gráficos devem ser apresentados conforme os princípios de representação visual de informação que estudamos em aula.

Contexto:

A população e a produção econômica é bastante concentrada espacialmente. Em particular, algumas poucas cidades concentram uma fração grande da renda e da produção. A região metropolitana de São Paulo é um exemplo dessa concentração, contendo 10% da população brasileira e sendo responsável por quase 17% do PIB do Brasil. A principal explicação para essa alta concentração é a existência de economias de aglomeração: mercados de trabalho mais populosos são também locais onde os trabalhadores são mais produtivos e, como consequência, tendem a ganhar salários maiores. Dessa forma, há um ciclo virtuoso: cidades maiores levam a salários maiores, que por sua vez atrai mais trabalhadores, que aumentam ainda mais os salários...

Há vários motivos para essas economias de aglomeração. Em particular, escala permite uma maior especialização e um funcionamento melhor dos mercados de trabalhos, em especial para trabalhadores especializados. Note que essas economias de aglomeração são estáticas, pois se referem a uma relação contemporânea (no mesmo ano) entre tamanho do mercado e salários.

Vários estudos empíricos confirmam a existência de economias de aglomeração estáticas, mostrando uma relação positiva entre o tamanho de um mercado de trabalho local e a produtividade de seus trabalhadores. Uma boa medida para o tamanho do mercado de trabalho é o número de trabalhadores (ou o seu logaritmo), enquanto uma boa medida para a produtividade dos trabalhadores é sua renda (lembre que, quando o mercado é competitivo, o salário tende a ser igual ao produto marginal do trabalho).

Roteiro:

Analise a existência de economias de aglomeração estáticas com base nos dados em anexo.

1. Há uma relação visível entre o logaritmo do emprego em uma microrregião e o logaritmo da renda média no ano 2000? Use um scatterplot. Documente o gráfico e escolha a sua formatação buscando clareza e a minimização do entulho.
2. Repita a mesma análise para o ano 2010. Como o gráfico mudou?
3. Calcule as correlações entre as duas variáveis para cada um dos anos. Interprete.
4. Uma dificuldade de identificação é que outras características de uma região, como acesso a infraestrutura de transporte e fontes de energia, podem estar positivamente correlacionadas tanto com o emprego e a renda. Dessa forma, a correlação entre cada uma dessas variáveis, em um dado ano, não necessariamente representa a relação causal que existiria devido às economias de aglomeração.

Uma solução para se lidar com esses problemas é “diferenciar” os dados. Ao invés de se estudar a relação entre os logaritmos de emprego e renda em 2000 ou em 2010, se examina a correlação entre as *diferenças* dessas variáveis entre 2000 e 2010. Logo, características de uma região que se mantenham constante no período não mais poderão viesar a relação. Construa essas diferenças para os logaritmos do emprego e da renda. Prepare um scatterplot e calcule a correlação. Como a sua interpretação mudou após esse novo exercício? Ela ainda é a mesma?

5. Uma forma de testar estatisticamente se um mercado de trabalho maior está associado a um salário maior consiste em dividir a amostra em dois grupos, um onde a população é maior e um onde a população é menor, e depois realizar um teste de diferenças de médias do logaritmo da renda entre os dois grupos.
 - a. Formalize e explique passo-a-passo esse teste. Em particular, explique o cutoff para população que você usaria.
 - b. Realize e reporte o resultado desse teste quando se usam dados do ano 2000.
 - c. Realize e reporte o resultado desse teste quando se usam dados do ano 2010.