



**UNIVERSIDADE FEDERAL DA PARAÍBA  
CENTRO DE CIÊNCIAS SOCIAIS APLICADAS  
DEPARTAMENTO DE ECONOMIA  
PROF. ANTÔNIO VINÍCIUS BARBOSA**

**Pedro Augusto A. F. de Melo**

**Gabriel Batista Pontes**

**PROJETO FINAL - ECONOMETRIA I**

**João Pessoa**

**23 de Outubro de 2024**

## NAVEGAÇÃO

Mercado de Trabalho.....	3
Taxa de Glicose e Incidência de Diabetes.....	5
Link para repositório.....	

## Análise do modelo salarial

O primeiro modelo deste projeto utiliza uma equação de Mincer para quantificar a relação existente entre o salário por hora do trabalhador com os anos de educação formal e o sexo do trabalhador. O modelo é descrito na equação abaixo:

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

1 -

Os resultados do modelo estimado estão descritos no quadro abaixo:

Estatística	Constante	educ	homem
Parâmetros	4.938	0.099	-0.033
Erro Padrão	0.045	0.003	0.024
P Values	0.000e+00	2.498e-182	1.604e-01
Intervalo de Confiança	[4.85, 5.025]	[0.093, 0.106]	[-0.079, 0.013]

2 -

A partir da análise dos resultados do modelo em questão, observa-se que o sinal do estimador *Beta1* associado à variável *educ* é positivo, o que evidencia uma relação direta entre a quantidade de anos formais de estudo e o salário de um trabalhador. Ainda, é importante ressaltar que, por causa da aplicação do *log* à variável dependente, o valor de *Beta1* representa a variação esperada relativa do salário do trabalhador (dividido por 100) quando se tem um ano a mais de educação formal. Sob esse contexto, o sinal esperado do estimador realmente era positivo, tendo em vista que espera-se que uma pessoa com uma formação mais completa tenha maiores salários.

3 -

Estatística	Mulher	Homem
Média Salarial (\$)	687.81	663.43

A equação que indica a diferença salarial média entre homens e mulheres com características idênticas é dada abaixo:

$$dif\_salarial\_media = - 0.033 \times Salario_i$$

Tomando os resultados como base, nota-se que o valor negativo de 0,033 para o estimador de Beta2 mostra o que a variável *dummy* homem altera o intercepto do modelo em aproximadamente 3,3% para baixo, sendo assim a *dummy* de intercepto do nosso modelo. Entretanto, quando o teste de hipótese é realizado, com a H0 afirmando que o Beta2 é nulo e a H1 afirmando que o Beta2 é não nulo, a um nível de significância alpha de 5%, o resultado do p-valor maior que alpha implica na não rejeição da H0. Ainda, há uma outra forma de se chegar à mesma conclusão, que é a apartir da análise do intervalo de confiança do Beta2, que inclui o valor 0, o que reforça o caráter não significativo do nosso estimador e, consequentemente mostra a falta de influência da variável sexo no salário do trabalhador, mesmo que o modelo e a análise das médias salariais mostrem o contrário.

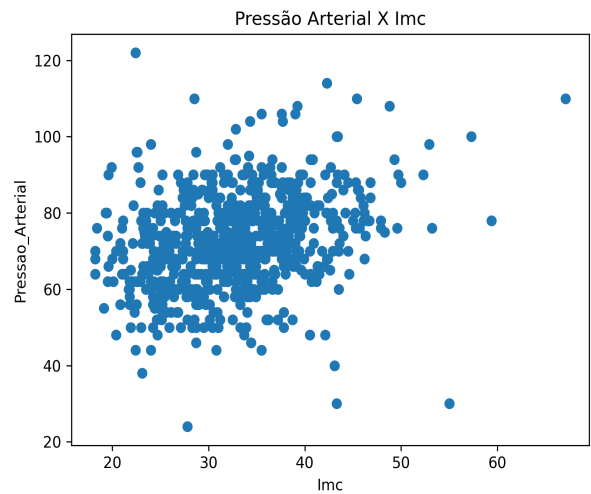
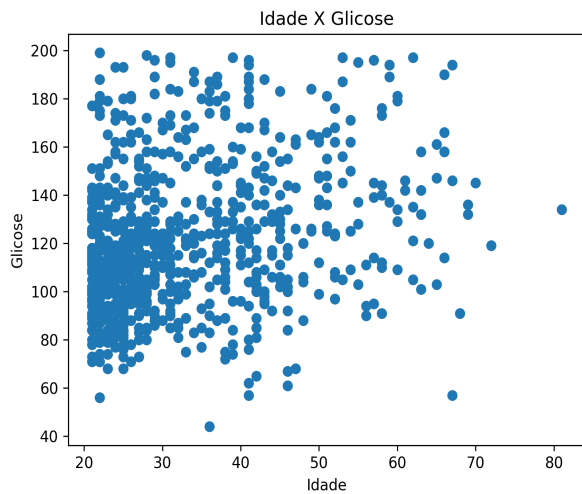
## Análise do modelo de diabetes

1 -

statistic	n_gravidez	glicose	pressao_arterial	insulina	imc	idade	diabetes
Média	3.85	120.89	69.11	79.8	31.99	33.24	0.35
Desvio Padrão	3.37	31.97	19.36	115.24	7.88	11.76	0.48

Nota: não faz sentido gerar média ou desvio padrão para a variável diabetes.

2 -



Acerca do gráfico que relaciona a idade do indivíduo com o nível de glicose, o conjunto de dados parece ter sido colhido de um grupo com grande concentração de idade entre 20 e 40 anos, o que pode trazer um pouco de viés à análise mas, em geral, o nível de glicose parece se elevar conforme a idade avança. Além disso, no segundo gráfico, este parece possuir uma concentração de observações no centro do plano, onde as pessoas possuem idade que varia de 20 a 40 anos e pressão de 40 a 100, mas sem uma evidência de relação entre as variáveis mais moderada.

3 -

### Idade x Glicose

Estatística	Constante	idade
Parâmetros	98.632	0.693
Erro Padrão	3.198	0.091
P Values	3.683e-136	6.208e-14
Intervalo de Confiança	[92.355, 104.91]	[0.515, 0.871]

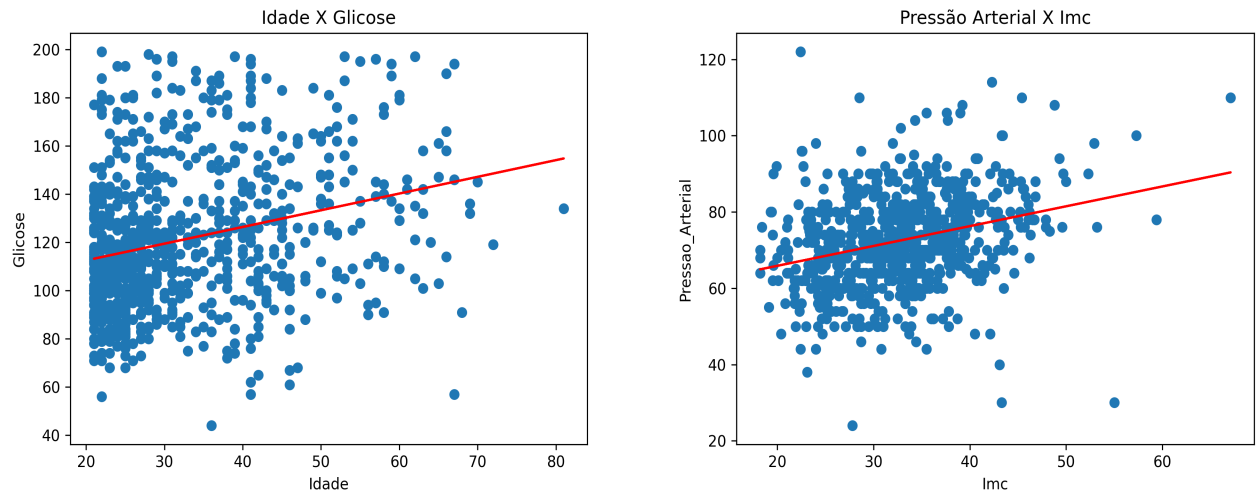
No primeiro modelo proposto, a constante (Beta0) mostra o nível de glicose quando o indivíduo tem 0 ano de idade. Além disso, o valor positivo de 0.693 para Beta1 mostra a relação positiva entre as variáveis, onde o nível de glicose varia em Beta1 ponto para cada ano a mais de idade do observado. Ainda sobre este modelo, observa-se que o p-valor dos estimadores apontam uma forte significância estatística, o que faz com que eles sejam considerados estimadores válidos para os parâmetros deste modelo, tendo em vista que podem ser considerados significantes a 1%.

### IMC x Pressão Arterial

Estatística	Constante	imc
Parâmetros	55.487	0.52
Erro Padrão	2.118	0.064
P Values	4.890e-107	1.630e-15
Intervalo de Confiança	[51.329, 59.645]	[0.395, 0.645]

Neste segundo modelo, a constante apresenta o valor esperado da pressão arterial para indivíduos com imc 0 (o que é impossível), enquanto o valor do coeficiente imc apresenta o valor de 0.52, mostrando que o nível da pressão arterial da mulher deve ser elevado em 0.52 unidade quando se tem uma variação de uma unidade no imc. Ambos os estimadores são estatisticamente significantes a 1%, validando o nosso modelo utilizado (analisando de forma isolada), podendo ser observado também na análise do intervalo de confiança a 95% de confiança, em que o 0 não está inserido nos intervalos.

4 -



5 & 6 -

Estatística	Constante	idade	d_gravidez
Parâmetros	97.318	0.73	-1.012
Erro Padrão	3.396	0.101	2.536
P Values	2.791e-123	1.027e-12	6.900e-01
Intervalo de Confiança	[90.652, 103.984]	[0.532, 0.928]	[-5.991, 3.967]

Neste último modelo construído, cujos resultados estão apresentados acima, a relação entre o nível de glicose e as variáveis idade e a dummy *d\_gravidez* é mensurada de forma linear e, a partir da análise do quadro acima, pode-se analisar a influência das variáveis independentes na taxa de glicose. Nesse contexto, observa-se que o valor da constante (97.318) mostra o valor do nível de glicose para mulheres que não tiveram gestações e com 0 ano. Ainda, conforme a idade é elevada em 1 unidade, ocorre uma variação no nível de glicose de 0.73 unidades, que é justamente o valor captado pelo estimador Beta1 associado à variável idade. Além disso, o estimador Beta2, coeficiente atrelado à variável *d\_gravidez*, com o valor negativo de -1.012 é uma variável *dummy* de intercepto, representando portanto a diferença média no nível de glicose de mulheres que já tiveram alguma gestação e das que não tiveram, a uma mesma idade.

Entretanto, ao observar o p-valor de cada um dos estimadores, nota-se que o Beta0 e o Beta1 são estatisticamente significantes, já que o seu p-valor é inferior ao alpha utilizado (5%). Já com relação ao Beta2, pode-se notar que o seu estimador não é estatisticamente significativo a 5%, o que nos leva a não rejeitar uma possível hipótese  $H_0$  de um teste de hipóteses, que afirma que o valor de Beta0 é nulo. Assim, tal variável *dummy* não deve ser validada neste modelo.