Estimação Bayesiana em Modelos Lineares com Aplicações em Economia

Trabalho de Conclusão de Curso

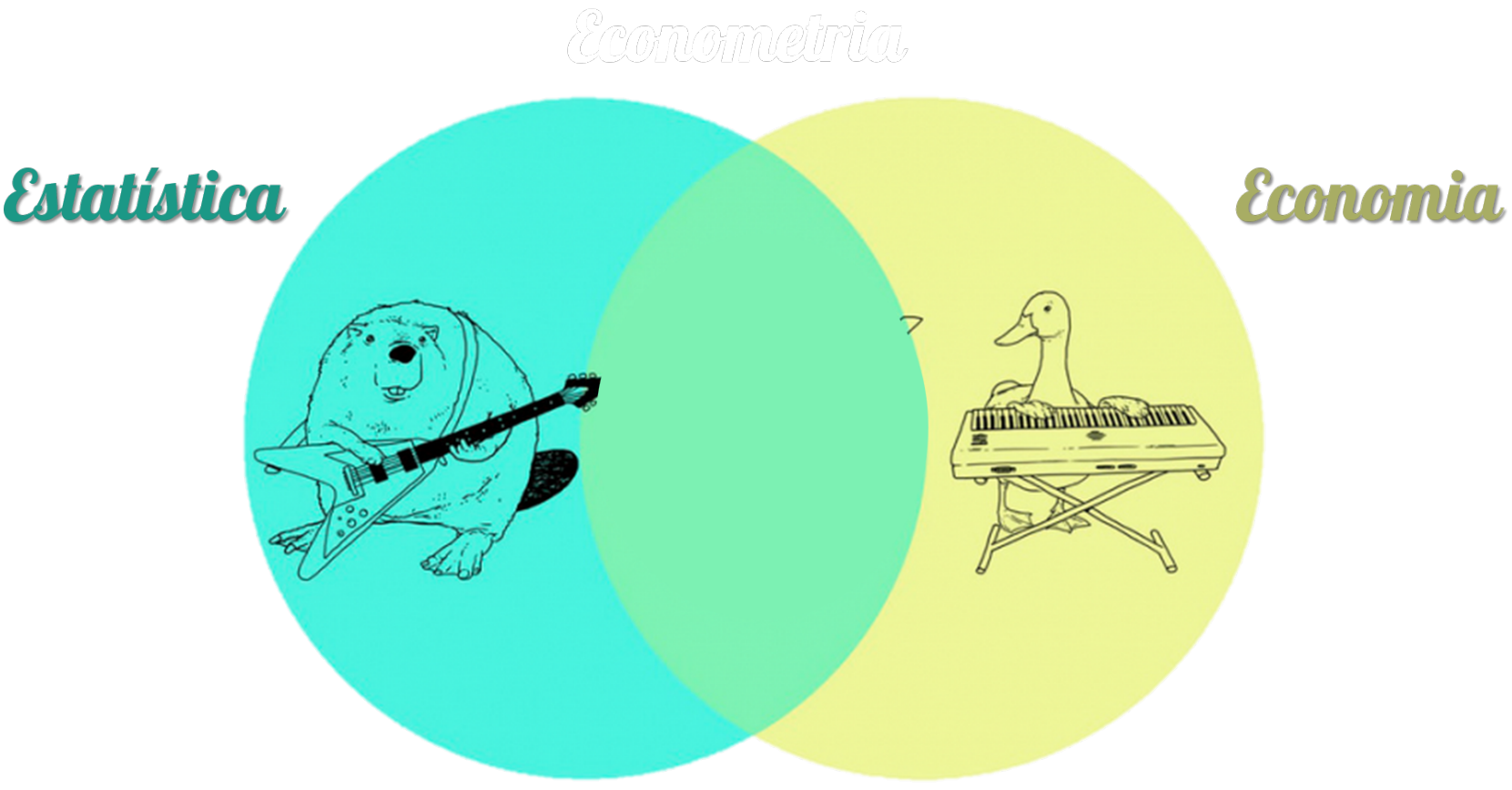
Aishameriane Schmidt
Orientador Prof. Fernando Pozzobon

UDESC - ESAG
Graduação em Ciências Econômicas

Fevereiro de 2018.

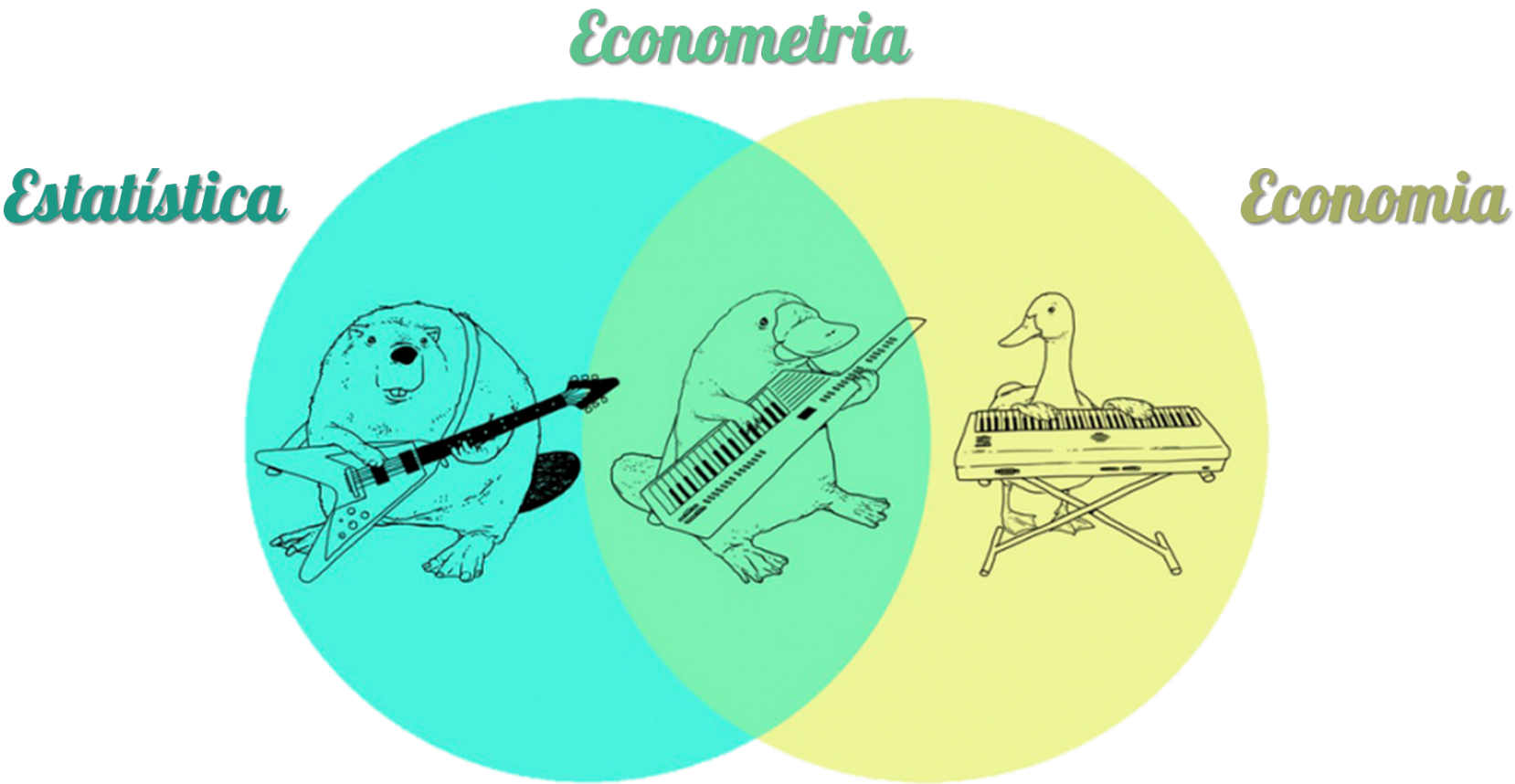
Econometria

Usando estatística para descrição e inferência de fenômenos econômicos



Econometria

Usando estatística para descrição e inferência de fenômenos econômicos



Roteiro

Estimação Bayesiana em Modelos Lineares com Aplicações em Economia

- Inferência Bayesiana
- O modelo de regressão linear normal
- Resultados analíticos
- Resultados de simulação
- Aplicação com dados do mercado de trabalho

Roteiro

Estimação Bayesiana em Modelos Lineares com Aplicações em Economia

- Inferência Bayesiana
- O modelo de regressão linear normal
- Resultados analíticos
- Resultados de simulação
- Aplicação com dados do mercado de trabalho

Inferência Bayesiana

O teorema de Bayes

Sejam A e B dois eventos dentro de um mesmo espaço de amostral com $\mathbb{P}(B) \neq 0$, então,

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} \tag{1}$$

Inferência Bayesiana

O teorema de Bayes

Sejam A e B dois eventos dentro de um mesmo espaço de amostral com $\mathbb{P}(B) \neq 0$, então,

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} \tag{1}$$

Se y representa um vetor de dados que segue uma densidade de probabilidade cujo vetor de parâmetros é θ, então, usando (1), podemos escrever:

$$\underbrace{\mathbb{P}(\theta|y)}_{\text{posteriori}} = \frac{\overbrace{\mathbb{P}(y|\theta)}^{\text{verossimilhança}} \overbrace{\mathbb{P}(\theta)}^{\text{priori}}}{\underbrace{\mathbb{P}(y)}_{\substack{\text{f.d.p. marginal} \\ \text{(constante)}}}} \tag{2}$$

Inferência Bayesiana

Um exemplo

Uma moeda é lançada 3 vezes e desejamos saber informações a respeito do parâmetro θ, que representa a probabilidade de sair cara.

Inferência Bayesiana

Um exemplo

Uma moeda é lançada 3 vezes e desejamos saber informações a respeito do parâmetro θ, que representa a probabilidade de sair cara.

- A verossimilhança é calculada a partir da distribuição de Binomial;

Inferência Bayesiana

Um exemplo

Uma moeda é lançada 3 vezes e desejamos saber informações a respeito do parâmetro θ, que representa a probabilidade de sair cara.

- A verossimilhança é calculada a partir da distribuição de Binomial;
- Na estimação clássica, sabemos que o EMV de θ é a média amostral
  - Ou seja, caso sejam observadas 3 caras, teremos

$$\hat{\theta}_{MV} = \frac{\sum_{i=1}^{3} y_i}{3} = \frac{3}{3} = 1$$

Inferência Bayesiana

Um exemplo

Uma moeda é lançada 3 vezes e desejamos saber informações a respeito do parâmetro θ, que representa a probabilidade de sair cara.

- A verossimilhança é calculada a partir da distribuição de Binomial;
- Na estimação clássica, sabemos que o EMV de θ é a média amostral
  - Ou seja, caso sejam observadas 3 caras, teremos

$$\hat{\theta}_{MV} = \frac{\sum_{i=1}^{3} y_i}{3} = \frac{3}{3} = 1$$

Será que temos informação suficiente para concluir que a probabilidade de sair cara nesta moeda é igual a 1?

Inferência Bayesiana

Um exemplo (contin.)

- Podemos considerar que θ é uma quantia aleatória e escolher uma densidade a priori:

Inferência Bayesiana

Um exemplo (contin.)

- Podemos considerar que θ é uma quantia aleatória e escolher uma densidade a priori:
  - Como θ é uma probabilidade, deve estar entre zero e um;

Inferência Bayesiana

Um exemplo (contin.)

- Podemos considerar que θ é uma quantia aleatória e escolher uma densidade a priori:
  - Como θ é uma probabilidade, deve estar entre zero e um;
  - Além disso, vamos supôr que θ assume *qualquer* valor neste intervalo;

Inferência Bayesiana

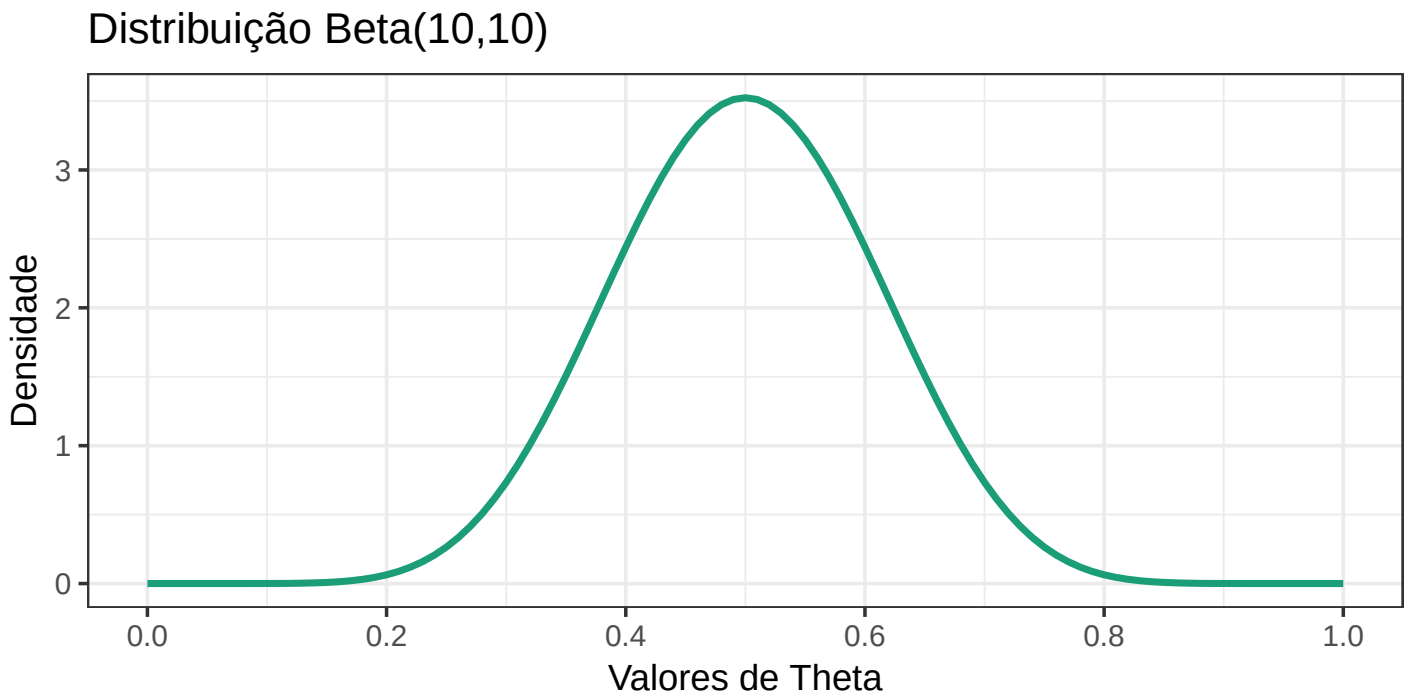
Um exemplo (contin.)

- Podemos considerar que θ é uma quantia aleatória e escolher uma densidade a priori:
  - Como θ é uma probabilidade, deve estar entre zero e um;
  - Além disso, vamos supôr que θ assume *qualquer* valor neste intervalo;
  - Vamos assumir a moeda é (possivelmente) honesta, isto é, acreditamos que θ tem uma maior "chance" de ser 50%.

Inferência Bayesiana

Um exemplo (contin.)

- Podemos considerar que θ é uma quantia aleatória e escolher uma densidade a priori:
  - Como θ é uma probabilidade, deve estar entre zero e um;
  - Além disso, vamos supôr que θ assume *qualquer* valor neste intervalo;
  - Vamos assumir a moeda é (possivelmente) honesta, isto é, acreditamos que θ tem uma maior "chance" de ser 50%.



Inferência Bayesiana

Um exemplo (contin.)

- A combinação de uma priori Beta($\underline{\alpha}$, $\underline{\beta}$) com uma verossimilhança que vem de uma densidade Binomial(n, θ) resulta em uma posteriori Beta com parâmetros $\bar{\alpha} = (\underline{\alpha} + \sum y_i)$ e $\bar{\beta} = (\underline{\beta} + n - \sum y_i)$.

Inferência Bayesiana

Um exemplo (contin.)

- A combinação de uma priori Beta($\underline{\alpha}$, $\underline{\beta}$) com uma verossimilhança que vem de uma densidade Binomial(n, θ) resulta em uma posteriori Beta com parâmetros $\bar{\alpha} = (\underline{\alpha} + \sum y_i)$ e $\bar{\beta} = (\underline{\beta} + n - \sum y_i)$.

- No exemplo das moedas, assumindo $\underline{\alpha} = \underline{\beta} = 10$, $n = 3$ e $\sum y_i = 3$, temos que a posteriori para θ é uma densidade Beta(13,10).

Inferência Bayesiana

Um exemplo (contin.)

- A combinação de uma priori Beta($\underline{\alpha}$, $\underline{\beta}$) com uma verossimilhança que vem de uma densidade Binomial(n, θ) resulta em uma posteriori Beta com parâmetros $\bar{\alpha} = (\underline{\alpha} + \sum y_i)$ e $\bar{\beta} = (\underline{\beta} + n - \sum y_i)$.

- No exemplo das moedas, assumindo $\underline{\alpha} = \underline{\beta} = 10$, $n = 3$ e $\sum y_i = 3$, temos que a posteriori para θ é uma densidade Beta(13,10).

- Consequentemente, $\mathbb{E}[\theta|y] = \frac{\bar{\alpha}}{\bar{\alpha}+\bar{\beta}} = \frac{13}{23} \approx 0.57$

Inferência Bayesiana

Um exemplo (contin.)

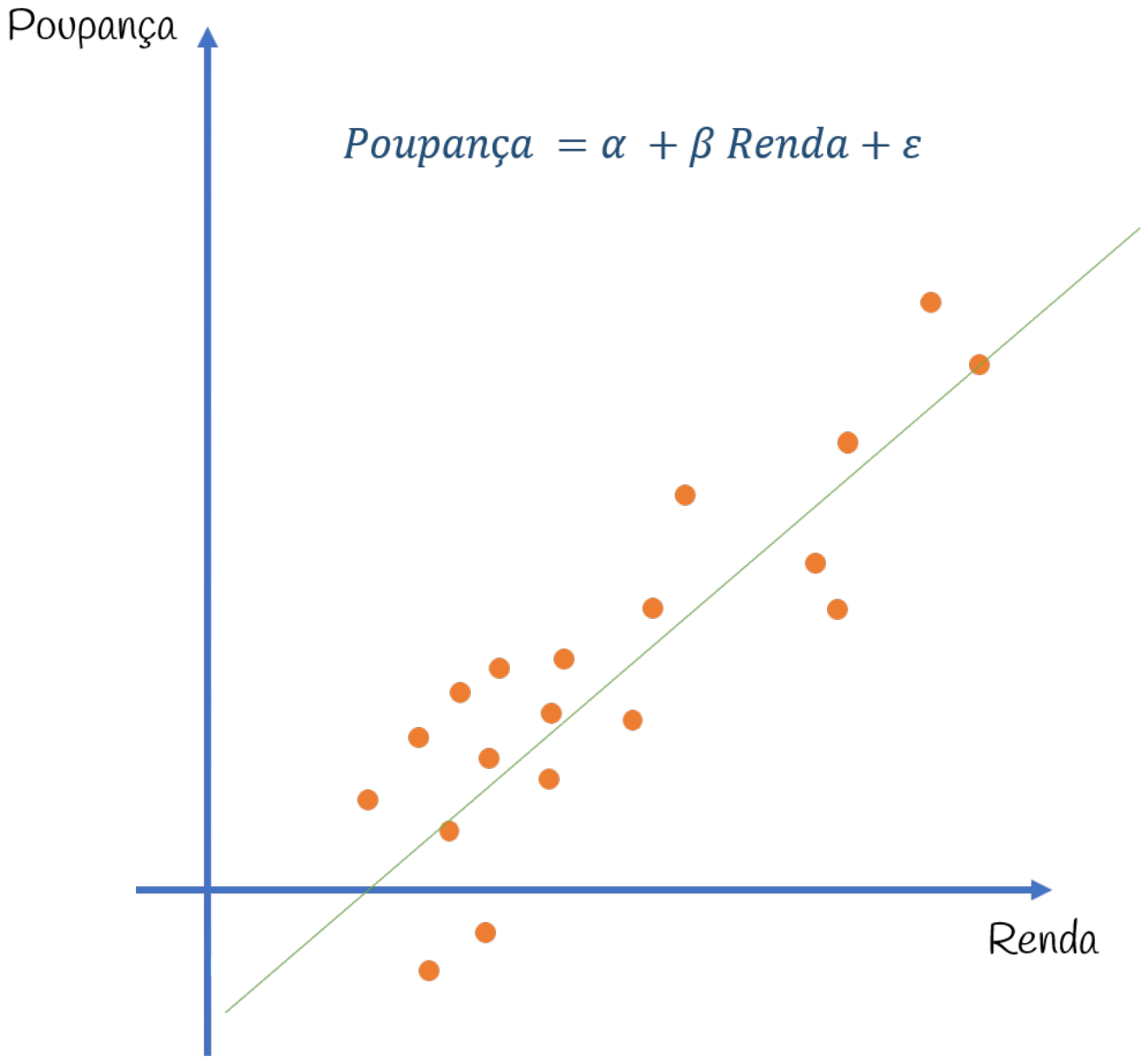
- A combinação de uma priori Beta($\underline{\alpha}$, $\underline{\beta}$) com uma verossimilhança que vem de uma densidade Binomial(n, θ) resulta em uma posteriori Beta com parâmetros $\bar{\alpha} = (\underline{\alpha} + \sum y_i)$ e $\bar{\beta} = (\underline{\beta} + n - \sum y_i)$.

- No exemplo das moedas, assumindo $\underline{\alpha} = \underline{\beta} = 10$, $n = 3$ e $\sum y_i = 3$, temos que a posteriori para θ é uma densidade Beta(13,10).

- Consequentemente, $\mathbb{E}[\theta|y] = \frac{\bar{\alpha}}{\bar{\alpha}+\bar{\beta}} = \frac{13}{23} \approx 0.57$
  - **Vantagens da abordagem bayesiana (neste exemplo)**: é menos suscetível ao excesso de ajuste que o EMV; podemos incorporar (de forma fácil) novos dados para atualização da priori quando lançarmos a moeda mais vezes.

O modelo de regressão linear Normal

Ideias básicas



Aplicação Econômica

Motivação e Hipótese

- O salário de um indivíduo é função de suas **características individuais** (experiência, escolaridade, produtividade, etc) e também de **fatores externos** (piso salarial, competitividade entre as empresas, dentre outros).

Aplicação Econômica

Motivação e Hipótese

- O salário de um indivíduo é função de suas **características individuais** (experiência, escolaridade, produtividade, etc) e também de **fatores externos** (piso salarial, competitividade entre as empresas, dentre outros).

Hipótese

Isoladas as características individuais, **não** deveria haver diferenciação de salário entre homens e mulheres que ocupam os mesmos tipos de cargos.

Aplicação Econômica

Revisão de Literatura

[Bonini and Pozzobon, 2016] utilizaram um modelo de regressão linear múltipla para investigar os salários e características individuais de trabalhadores da área de tecnologia da informação (TI) e trabalhadores da indústria nos três estados da Região Sul do Brasil no ano de 2011.

Aplicação Econômica

Revisão de Literatura

[Bonini and Pozzobon, 2016] utilizaram um modelo de regressão linear múltipla para investigar os salários e características individuais de trabalhadores da área de tecnologia da informação (TI) e trabalhadores da indústria nos três estados da Região Sul do Brasil no ano de 2011.

Controlados os demais fatores, seus resultados mostraram que **as mulheres em Santa Catarina que trabalhavam área de TI recebiam, em média, um salário 15% inferior ao dos homens**.

Aplicação Econômica

Modelo Econométrico

$$\begin{aligned}\ln(\text{salário}) = {} & \beta_0 + \beta_1 \cdot \text{Sup. Inc.} + \beta_2 \cdot \text{Sup. Comp.} + \beta_3 \cdot \text{Pós-Grad.} \\ & + \beta_4 \cdot \text{Idade} + \beta_5 \cdot \text{Idade}^2 + \beta_6 \cdot \text{T. Emp.} + \beta_7 \cdot \text{Gênero} + \varepsilon \end{aligned} \quad (3)$$

Aplicação Econômica

Base de dados

- Dados de salário, idade, escolaridade, tempo de emprego e gênero de 4318 trabalhadores retirados da RAIS;
- Setor de tecnologia da informação do Estado de Santa Catarina;
- Ano de 2011.

Aplicação Econômica

Decrição da amostra

Tabela 3 – Estatísticas descritivas das variáveis quantitativas do modelo

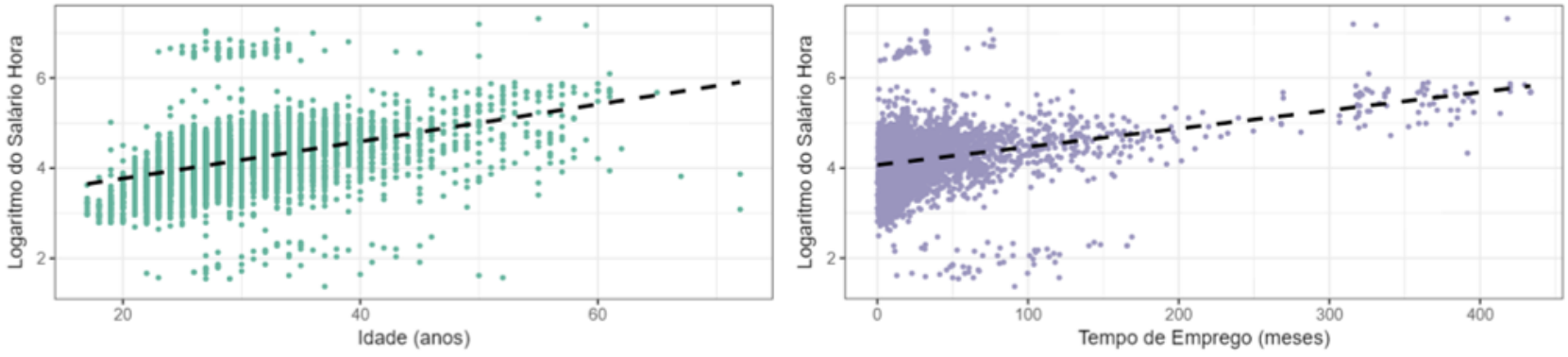
	ln(Salário-Hora)*	Tempo de Emprego†	Idade**	$(\text{Idade})^2$
N^o de Observações	4318	4318	4318	4318
Mínimo	1.38	0.10	17.00	289.00
1^o quartil	3.78	9.70	26.00	676.00
Mediana	4.22	16.90	29.00	841.00
Média	4.21	34.59	30.58	989.10
3^o quartil	4.60	37.10	34.00	1156
Máximo	7.32	433.40	72.00	5184
Desvio Padrão	0.69	54.26	7.35	527.32

* O salário-hora é medido em reais ($R\$$) de 2011 e ln(Salário-Hora) denota seu logaritmo na base neperiana.
† Tempo de emprego é medido em meses.
** Idade é medida em anos e, consequentemente, idade ao quadrado, denotada por $(\text{Idade})^2$, tem como unidade anos ao quadrado.
Fonte – Elaboração própria utilizando dados da RAIS de trabalhadores de setores de TI no estado de Santa Catarina para o ano de 2011.

Aplicação Econômica

Decrição da amostra

Figura 5 – Gráficos de dispersão das variáveis explicativas quantitativas contra a variável explicada no modelo



(a) Distribuição da idade e logaritmo do salário hora

(b) Distribuição do tempo de emprego e logaritmo do salário hora

Fonte – Elaboração própria com base nos dados da RAIS.

Aplicação Econômica

Decrição da amostra

Tabela 5 – Proporções nas categorias das variáveis categóricas incluídas no modelo

	Sim		Não	
	n	**%**	**n**	**%**
Homem*	3285	76,08	1033	23,92
Ensino superior incompleto†	697	16,14	3621	83,86
Ensino superior completo†	3201	74,13	1117	25,87
Pós graduação†	29	0,0067	4289	99,33

* A categoria "Homem" é considerada a categoria base para a variável do modelo "Gênero".

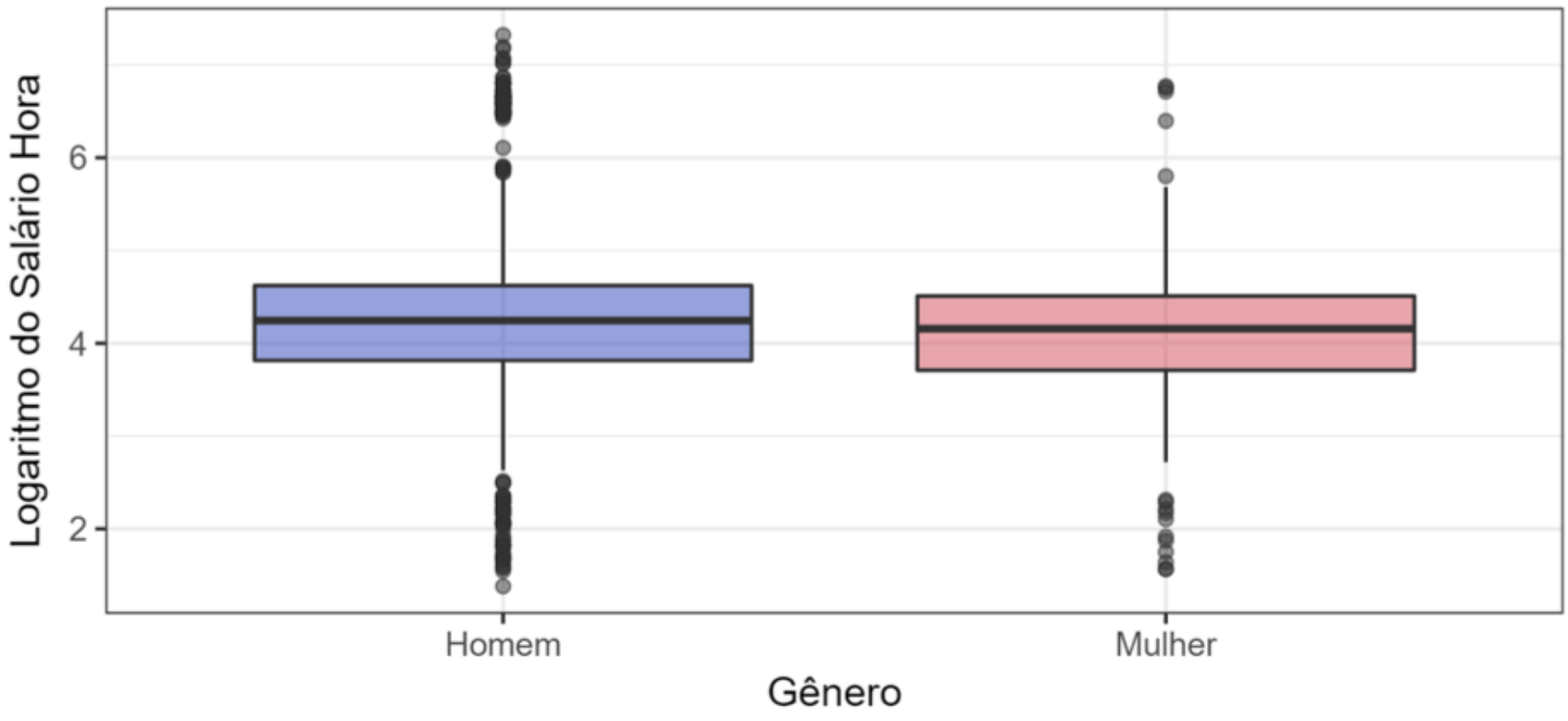
† As variáveis de escolaridade representam o grau de escolaridade máximo do indivíduo, isto é, uma pessoa que tenha cursado até o primeiro ano da faculdade terá "Sim" para ensino superior incompleto e "Não" para ensino superior completo e pós graduação. A categoria base da variável escolaridade é "Ensino Médio".

Fonte – Elaboração própria utilizando dados da RAIS de trabalhadores de setores de TI no estado de Santa Catarina para o ano de 2011.

Aplicação Econômica

Decrição da amostra

Figura 6 – Boxplot do logaritmo do salário hora estratificado por gênero para trabalhadores de TI no Estado de Santa Catarina no ano de 2011



Fonte – Elaboração própria com base nos dados da RAIS.

Aplicação Econômica

Coeficientes estimados

Tabela 6 – Valores estimados usando MQO e o método bayesiano com priori normal gama conjugada

Parâmetro	Priori	MQO	D.P. MQO	Posteriori	D.P. Posteriori	IC95%	$P(\bar{\beta}_j \geq 0)$
Constante	2.0712	1.49258	0.1238	1.49286	0.1249	[1.2481 ; 1.7376]	>0.9999
Sup. Inc	0.1227	0.09413	0.0368	0.09411	0.0371	[0.0213 ; 0.1669]	0.99438
Sup. Comp.	0.4487	0.3738	0.0319	0.37381	0.0321	[0.3108 ; 0.4368]	>0.9999
Pós	1.0708	1.07101	0.1124	1.07102	0.1134	[0.8487 ; 1.2934]	>0.9999
Idade	0.1285	0.11962	0.0072	0.11961	0.0073	[0.1053 ; 0.1339]	>0.9999
Idade^2	-0.0014	-0.00132	0.0001	-0.00132	0.0001	[-0.0015 ; -0.0011]	<0.0001
Temp. Emp.	0.0027	0.00252	0.0002	0.00252	0.0002	[0.0021 ; 0.0029]	>0.9999
Gênero	-0.1537	-0.13239	0.0208	-0.13239	0.0210	[-0.1736 ; -0.0912]	<0.0001

Fonte – Elaboração própria.

Conclusões

- Foi apresentado o modelo de regressão linear normal com priori conjugada natural;
  - Os principais resultados analíticos foram demonstrados;

Conclusões

- Foi apresentado o modelo de regressão linear normal com priori conjugada natural;
  - Os principais resultados analíticos foram demonstrados;
  - Realizaram-se experimentos com valores simulados para verificar a sensibilidade das estimativas frente a mudanças na priori, variáveis explicativas, especificação do modelo e presença de erro de medida, comparando as estimativas bayesianas com o estimador de MQO;

Conclusões

- Foi apresentado o modelo de regressão linear normal com priori conjugada natural;
  - Os principais resultados analíticos foram demonstrados;
  - Realizaram-se experimentos com valores simulados para verificar a sensibilidade das estimativas frente a mudanças na priori, variáveis explicativas, especificação do modelo e presença de erro de medida, comparando as estimativas bayesianas com o estimador de MQO;
  - Utilizando dados do trabalho de [Bonini and Pozzobon, 2016], replicou-se o artigo original.

Conclusões

- Métodos bayesianos em economia são uma ferramenta poderosa que possibilitaram avanços importantes em pesquisas acadêmicas, Bancos Centrais, etc;

Conclusões

- Métodos bayesianos em economia são uma ferramenta poderosa que possibilitaram avanços importantes em pesquisas acadêmicas, Bancos Centrais, etc;
- São uma alternativa aos métodos clássicos, especialmente em aplicações de alta dimensão onde o tratamento da função de verossimilhança é difícil (ou impossível);

Conclusões

- Métodos bayesianos em economia são uma ferramenta poderosa que possibilitaram avanços importantes em pesquisas acadêmicas, Bancos Centrais, etc;
- São uma alternativa aos métodos clássicos, especialmente em aplicações de alta dimensão onde o tratamento da função de verossimilhança é difícil (ou impossível);
- Possibilitam incorporar conhecimento a priori;

Conclusões

- Métodos bayesianos em economia são uma ferramenta poderosa que possibilitaram avanços importantes em pesquisas acadêmicas, Bancos Centrais, etc;
- São uma alternativa aos métodos clássicos, especialmente em aplicações de alta dimensão onde o tratamento da função de verossimilhança é difícil (ou impossível);
- Possibilitam incorporar conhecimento a priori;
- Não são tão sensíveis ao tamanho da amostra como métodos clássicos;

Conclusões

- Métodos bayesianos em economia são uma ferramenta poderosa que possibilitaram avanços importantes em pesquisas acadêmicas, Bancos Centrais, etc;
- São uma alternativa aos métodos clássicos, especialmente em aplicações de alta dimensão onde o tratamento da função de verossimilhança é difícil (ou impossível);
- Possibilitam incorporar conhecimento a priori;
- Não são tão sensíveis ao tamanho da amostra como métodos clássicos;
- Permitem atualização dos modelos de forma mais direta e iterativa.

Estimação Bayesiana em Modelos Lineares com Aplicações em Economia

Trabalho de Conclusão de Curso

Aishameriane Schmidt
Orientador Prof. Fernando Pozzobon

UDESC - ESAG
Graduação em Ciências Econômicas

Fevereiro de 2018.

aishameriane@gmail.com

Referências

Bonini, P. and Pozzobon, F. (2016).
Discriminação salarial feminina e o prêmio salarial de ti na indústria.
Análise Econômica (UFRGS), 34:193–223.