



SPOLM 2007

ISSN 2175-6295

Rio de Janeiro- Brasil, 08 e 09 novembro de 2007.

# UMA ANÁLISE BAYESIANA DO PERFIL COR/RAÇA DOS INDIVÍDUOS QUE TEM CURSO SUPERIOR NO BRASIL

**Giovani Glaucio de Oliveira Costa**

Universidade Federal Rural do Rio de Janeiro

Instituto Multidisciplinar

Rua Professor Paris S/N. Centro. Nova Iguaçu. Rio de Janeiro. Brasil

giovani@ufrj.br

## Resumo

*O Brasil ainda é um país em que as desigualdades de oportunidades entre brancos e negros/pardos é muito grande. No campo educacional, isto fica ainda mais evidente quando se analisa quem tem curso superior. É postulável a hipótese de que a maior parte dos brasileiros que já tenham diploma de curso universitário seja da raça branca. Esta informação a priori, anterior ou externa em relação à amostra ou experiência, é demasiado importante para ser ignorada. Para tratar adequadamente e formalmente a referida informação este artigo introduz uma análise bayesiana ao problema, onde uma distribuição a posteriori é calculada, considerando a informação prévia assumida e o particular resultado de uma experiência aleatória, através da observação da amostra da PNAD 2002, no Teorema de Bayes. A revisão das credibilidades iniciais pelo resultado da amostra permitiu-se fazer análises que, no presente caso, consistiu em aumentar consideravelmente à credibilidade de ser branca a cor/raça mais provável dos que tem curso superior no Brasil. A distribuição a posteriori se constituiu na estrutura em que se assentou todo o trabalho analítico deste estudo.*

**Palavras-chaves:** brasileiros, brancos, negros/pardos, probabilidades a priori, probabilidades a posteriori, análise bayesiana.

## 1-Introdução

Segundo dados da Pesquisa Nacional da Amostra Domiciliar(PNAD,2002) apesar de ter havido o crescimento na média de anos de estudo da população de 15 anos ou mais (aproximadamente 1,5 anos) ao longo da década 1993-2004, o diferencial entre negros e brancos caiu muito pouco neste período. Se em 1993 esse diferencial era de 2,1 anos a favor dos brancos, em 2004 ele cai apenas para 1,9.

Essa diferença fica amplificada se comparamos a média de anos de estudo da população negra da região Nordeste – 5,1 em 2004 – com o mesmo indicador para a população branca da região Sudeste – 8,2 anos. Isto é, enquanto os negros nordestinos

estão concluindo somente a primeira etapa do ensino fundamental (o antigo primário), os brancos da região Sudeste, na média, concluem os 8 anos desse nível de ensino.

As diferenças raciais tendem a se perpetuar, caso não sejam empreendidas ações voltadas a reverter o atual quadro, no qual as diferenças entre brancos e negros impressionam: a média de 7,7 anos de estudo entre os brasileiros brancos se contrapõe a 5,8 anos entre os brasileiros negros.

Com o objetivo de alertar as autoridades competentes sobre o problema em foco, este trabalho objetiva determinar as probabilidades condicionais de que um brasileiro tendo curso superior seja de origem racial branca e de origem racial negra/parda e também fazer uma análise bayesiana do problema.

## **2-Amostragem**

O presente trabalho utiliza uma amostra de trabalhadores com idade entre 25 e 64 anos, residentes nas áreas urbanas do Brasil, extraída da PNAD (Pesquisa Nacional por Amostra de Domicílios) de 2002. Os dados correspondem a uma amostra aleatória de 10% das observações da PNAD.

A PNAD é uma pesquisa anual feita pelo IBGE representativa de toda a população brasileira. Em cada ano são entrevistados em torno de 330.000 indivíduos em cerca de 100.000 domicílios. A amostra utilizada envolveu 10.014 indivíduos da amostra original da PNAD.

Para realização do presente estudos são consideradas as variáveis: Superior=1, se a pessoa pelo menos completou o curso superior e 0 em caso contrário e branco=1, se a pessoa é branca e 0 se negro/pardo.

## **3-Cálculo das Probabilidades de Bayes e Formulação das Hipóteses a Priori**

No Brasil, uma constatação evidente: a maior parte dos brasileiros que tem curso superior são brancos.

Esta informação é importante como indicador das desigualdades sociais imposta pela origem racial das pessoas, estando latente também no problema a questão sócio-econômica, uma vez que o grau de escolaridade das pessoas é um fator de ascensão social e econômica.

O Brasil já mostrou nos últimos anos um avanço para reversão deste quadro, mas ainda é muito cedo para que se constate que no país exista igualdade de condições de estudos, em consequência de ascensão social e econômica, entre brancos e negros/pardos.

Neste artigo se considerou a origem racial dos brasileiros de forma bem sucinta:

- Branco
- Negro/pardo

Os brasileiros possuidores de diploma de curso superior ou são da raça branca ou da raça negra/parda. O percentual de outras raças foi tomado como insignificante. A idéia deste trabalho é permitir o raciocínio bayesiano e, portanto, é oportuno formular hipóteses a priori quanto às probabilidades prováveis de brasileiros com curso superior em ambas as categorias: da raça branca ou não da raça branca.

O quadro 1 apresenta as informações assumidas para as distribuições de probabilidades dos brancos e negros/pardos (52% e 48%, respectivamente) e para as

distribuições de probabilidades condicionais de brancos e negros que tem curso superior no Brasil( 20% e 10%, respectivamente).

**Quadro 1:** Hipóteses quanto à Distribuição das Probabilidades do Problema

Branco <b>52%</b>	Negro/pardo <b>48%</b>
Curso Superior <b>20%</b>	Curso Superior <b>10%</b>
Caso Contrário <b>80%</b>	Caso Contrário <b>90%</b>

Seja a seleção, aleatória, de um brasileiro. Admita-se que a pessoa tenha curso superior. Uma pergunta que se pode fazer é se é possível determinar a probabilidade de que o brasileiro, tendo curso superior, é de origem racial branca ou é de origem negra/parda. Propõe-se responder afirmativamente a esta pergunta.

Tem-se:

$P(B)=0,52 \rightarrow$  probabilidade de que o brasileiro seja branco

$P(NP)=0,48 \rightarrow$  probabilidade de que o brasileiro seja negro/pardo.

Pelo teorema do cálculo das probabilidades, verifica-se:

$$P(B \cap S) = P(B) \cdot P(S/B) \quad (1)$$

$$P(NP \cap S) = P(NP) \cdot P(S/NP) \quad (2)$$

$(S/B)$  e  $(S/NP)$  são informações úteis.

A probabilidade de o brasileiro selecionado ter curso superior é igual à soma das probabilidades de que o brasileiro seja branco com a probabilidade de que o brasileiro seja negro/pardo, como mostra a expressão (3).

$$P(S) = P(B \cap S) + P(NP \cap S) \quad (3)$$

As probabilidades descritas nas expressões (1) e (2), que resultam na probabilidade da expressão (3), não são conhecidas sob forma direta, porém podem ser calculadas em função das probabilidades condicionais conhecidas:

Já se sabe que:

$$P(B \cap S) = P(B) \cdot P(S/B)$$

$$P(NP \cap S) = P(NP) \cdot P(S/NP)$$

Então:

$$P(B/S) = [P(B) \cdot P(S/B)] / P(S) \quad (4)$$

$$P(NP/S) = [P(NP) \cdot P(S/NP)] / P(S) \quad (5)$$

As fórmulas (4) e (5) resultam portanto:

$$P(B/S) = [P(B) \cdot P(S/B)] / [P(B) \cdot P(S/B) + P(NP) \cdot P(S/NP)] \quad (6)$$

$$P(NP/S) = [P(NP) \cdot P(S/NP)] / [P(B) \cdot P(S/B) + P(NP) \cdot P(S/NP)] \quad (7)$$

Os modelos (6) e (7) constituem no Teorema de Bayes. Substituindo os valores assumidos do problema nas expressões (6) e (7) tem-se:

$$P(B/S) = [0,52 \cdot 0,20] / [0,52 \cdot 0,20] + [0,48 \cdot 0,10] = \mathbf{0,68}$$

$$P(NP/S) = [0,48 \cdot 0,10] / [0,52 \cdot 0,20] + [0,48 \cdot 0,10] = \mathbf{0,32}$$

Até agora se aplicou simplesmente o Teorema de Bayes e utilizando-se de dados que são meramente hipóteses, idéias pré-concebidas. Na próxima seção, se irá mostrar como utilizar a informação amostral disponível para estimar as proporções condicionais estabelecidas no problema.

#### 4-Uma Análise Bayesiana do Problema:

Lembrando que se formularam as hipóteses iniciais de que 52% dos brasileiros são da raça branca, contra 48% da raça negra/parda. O objetivo atual é modificar estas probabilidades para que reflitam nas seguintes condicionais:

- ✓ *Probabilidade de que um indivíduo, sendo de curso superior, seja da cor/raça branca;*
- ✓ *Probabilidade de que um indivíduo, sendo de curso superior, seja da cor/raça negra/parda.*

Aceitam-se os dados descritos no Quadro 2, como probabilidades a priori.

## Quadro 2- Hipóteses ou Probabilidades a Priori

Sucessos	Probabilidades a Priori
B	0,52
NP	0,48
$\Sigma$	1,00

Coerente com a teoria bayesiana, se proporá modificar, isto é, corrigir as hipóteses a priori, com vistas à informação amostral, fornecida pela PNAD.

Com os dados fornecidos pela PNAD, pode-se fazer uma análise bayesiana do problema: é o que revela o Quadro 3:

## Quadro 3- Hipóteses (Probabilidades a Priori) e Fator de Verossimilhança (Probabilidades Condicionais)

Sucessos	Probabilidades a Priori	Fator de Verossimilhança (Probabilidades Condicionais)
B	0,52	0,16=P(S/B)
NP	0,48	0,06=P(S/NP)

A partir do Quadro 3, pode-se obter o Quadro 4, que apresenta o cálculo para as probabilidades a posteriori.

## Quadro 4- Quadro de Cálculo das Probabilidades a Posteriori

Sucessos	Hipóteses quanto a Distribuição Racial (1)	Verossimilhança (2)	Prob. Conjunta (3)=(1) . (2)	Probabilidades a Posteriori (4)=[(1).(2)] / P(S)
B	0,52	0,16	0,08	<b>0,73</b>
NP	0,48	0,06	0,03	<b>0,27</b>
$\Sigma$	1,00	-	<b>0,11=P(S)</b>	<b>1,00</b>

A revisão das credibilidades iniciais, descrita no Quadro 4, pelo resultado da amostra particular, permite se fazer inferências que no presente caso, consiste em aumentar substancialmente a credibilidade do branco, como cor/raça mais provável dos que tem curso superior e reduzir consideravelmente à credibilidade de ser negra/parda a cor dos que tem diploma universitário no país.

Então, dado que um brasileiro tenha curso superior, é bem mais provável que ele seja branco(73%).

## 5-Conclusão

Este trabalho objetivou estimar a porcentagem dos brasileiros com curso universitário em ambas as categorias de cor/raça: branca ou negra/parda, realizando para tal uma análise bayesiana do problema.

Observando os resultados obtidos, conclui-se que a influência da Pesquisa Nacional por Amostra Domiciliar do IBGE, a PNAD, realizada sob as hipóteses iniciais, aumenta para 73% o percentual de brasileiros de curso superior que são brancos e reduz para 27% o percentual de brasileiros de curso superior que são negros/pardos no Brasil. A credibilidade da cor branca como o perfil mais provável dos que tem curso superior é muito alta.

A estimativa bayesiana do percentual de pessoas, que já tenham curso superior completo, serem da raça branca é de 73% e a estimativa do percentual de pessoas, que já tenham curso superior completo, serem da raça negra/parda é de 27%.

Mais uma vez a análise científica evidencia que a desigualdade de condições de ascensão social ainda persiste no país entre brancos e negros/pardos. Os brancos são ainda a grande maioria dos que tem curso superior no Brasil e conseqüentemente são os que mais tem oportunidades de ter uma vida melhor, em condições sócio-econômicas mais dignas.

Este quadro não se resolve com quotas em reserva nas universidades públicas, mas com o fomento de políticas públicas efetivas e sistemáticas que possibilitem o negro/pardo ter acesso ao curso superior de maneira natural e por real mérito intelectual. Somente assim se verá no Brasil um equilíbrio na proporção de cor/raça dos que tem ensino superior.

## 6-Bibliografia

1. **DeGroot, M.H. & Schervish, M.J.**-2002-Probability and Statistics, (3 Edição), Addison-//weley, New York.
2. **Cox,D.R. e Hinkley,D.V.** -1974- Theoretical Statistics, Chapman Hall, London.
3. **Davison,A.C.**-2003- Statistical Models, Cambridge University Press, Cambridge.
4. **James, B.R.** -1981- Probabilidade: UM Curso em Nível Intermediário, IMPA. Projeto Euclides.
5. **Serra Costa, José de Jesus da.** Teoria da Decisão: Um Enfoque Objetivo. Editora Rio.1977.
6. **Lee, P. M.**-1989-Bayesian Statistics: An Introduction. Oxford University Press. Oxford University. London.
7. **Migon,H.S. e Gamerman,D.**-1999-Statistical Inference : Na Integrated Approach. Arnold, London.
8. **Paulino,C.D. , Turkman, M.A. e Murteira,B.**-2003- Estatística Bayesiana.Fundação Calouste Gurberkian, Lisboa.