



Análise - Dados pré/pós (FGV)

Coordenação de Pesquisa e Avaliação

Fábio Rocha - fabio.vianna@institutounibanco.org.br

Sumário

Introdução	3
Vizualização das variáveis (pré)	3
Vizualização das variáveis (pós)	3
Descritivas da nota (pré)	4
Descritivas da nota (pós)	4
Teste de média	4
Cruzamento das bases a partir do Id	4
Delta (variação da nota) dos participantes	5
Análises da variação antes e depois	5
Pré treinamento	5
Pós treinamento	5
Análises das perguntas	6
Perguntas a serem respondidas:	6
Análise das variáveis do conjunto de dados (2)	6
Formato das perguntas do teste	8
Seleção dos dados pré teste	9

Introdução

Trata-se de um exercício preliminar exploratório não exaustivo de dados produzidos por participantes de um treinamento ao realizarem uma prova. O exercício apesar de produzido por um analista da Coordenação de Pesquisa e Avaliação (CPA) do Instituto Unibanco, contou com contribuições técnicas de outro analista, Fernando Marques, também da CPA.

Seis conjuntos de dados foram compartilhados.

- (1) Desempenho dos usuários nas provas antes e depois (dois conjuntos de dados)
- (2) Marcação em cada alternativa em questão antes e depois (dois conjuntos de dados)
- (3) Pontuação global em cada questão antes e depois (dois conjuntos de dados)

Por ora, para as perguntas propostas neste exercício, apenas dois bancos foram utilizados, (1) e (2)

O script da análise está disponível em []

Vizualização das variáveis (pré)

O conjunto de dados (1) contém 9 variáveis (colunas) e 57 observações, como pode ser visto a seguir. Como estamos interessados inicialmente apenas no desempenho médio antes e depois do treinamento, variáveis como nome, desvio padrão e outras não serão consideradas. Abaixo as cinco primeiras observações do conjunto de dados (1) pré.

##	ID.Definido.da.Organização	Nome.de.Usuário	Nome	Sobrenome	Pontuação	De	X.
## 1	A58470218	A58470218	Alysson	Aguiar	7	10	70%
## 2	A58455434	A58455434	ILSON	ALVES	7	10	70%
## 3	A58455832	A58455832	Sarah	Arruda	9	10	90%
## 4	A58471990	A58471990	Janaína	Audino	7	10	70%
## 5	A58455432	A58455432	Genilma	Barros	8	10	80%

##	Média.da.Classe	Desvio.Padrão.da.Classe
## 1	83,51%	12,77%
## 2	83,51%	12,77%
## 3	83,51%	12,77%
## 4	83,51%	12,77%
## 5	83,51%	12,77%

Vizualização das variáveis (pós)

As cinco primeiras observações dos mesmos dados, só que para prova após o treinamento.

##	ID.Definido.da.Organização	Nome.de.Usuário	Nome	Sobrenome	Pontuação	De
## 1	A58470218	A58470218	Alysson	Aguiar	10	10
## 2	A58455434	A58455434	ILSON	ALVES	7	10
## 3	A58455832	A58455832	Sarah	Arruda	10	10
## 4	A58455432	A58455432	Genilma	Barros	9	10
## 5	A58456808	A58456808	Paulo	Borges	10	10

##	X.	Média.da.Classe	Desvio.Padrão.da.Classe
## 1	100%	89,44%	9,11%
## 2	70%	89,44%	9,11%

## 3	100%	89,44%	9,11%
## 4	90%	89,44%	9,11%
## 5	100%	89,44%	9,11%

Descritivas da nota (pré)

Em termos gerais o desempenho na prova antes do treinamento apresenta a distribuição da tabela. Note que a menor nota foi 4 pontos e que média e mediana se aproximam. Ou seja, 50% dos participantes performaram com nota 8 ou mais.

min	media	Q1	Q2	Q3	max	desvio_padrao
4	8.35	7	8	9	10	1.29

Descritivas da nota (pós)

Já após o treinamento a distribuição melhora em todos os parâmetros. A média sobe 0.59, a menor nota não mais é 4 e sim 6 e agora 50% dos participantes obtiveram nota 9 ou mais.

min	media	Q1	Q2	Q3	max	desvio_padrao
6	8.94	8	9	10	10	0.92

Teste de média

A fim de verificar se essa média é estatisticamente relevante a um dado nível de significância, produzimos um teste de médias a partir de OLS (que produz o mesmo resultado em um teste t padrão). O resultado é que a melhora de 0.5 pontos é significativa a 1%. Em termos percentuais (mesmo procedimento mas com a variável nota na sua forma logarítmica) indica uma melhora de 7,6% na nota após o treinamento.

Table 3: Teste de média pré-pós

	Dependent variable:	
	pontuacao (1)	log(pontuacao) (2)
tratamento	0.594*** (0.214)	0.076*** (0.027)
Constant	8.351*** (0.149)	2.109*** (0.019)
Observations	111	111
R ²	0.066	0.067
Adjusted R ²	0.058	0.058
Residual Std. Error (df = 109)	1.124	0.144
F Statistic (df = 1; 109)	7.727***	7.805***
Note:	*p<0.1; **p<0.05; ***p<0.01	

Cruzamento das bases a partir do Id

A partir desse procedimento Dos participantes totais na fase pre treinamento, 3 não realizaram a prova pós treinamento.

Delta (variação da nota) dos participantes

Ou seja, como já verificado no teste de média, a melhora foi de 0.55 (0.5 em 10). Também foi verificado situação em que o usuário obteve piora no desempenho, vide o valor mínimo -2.

Dito de outro modo, 50% dos usuário variaram de uma nota negativa a uma nota igual a anterior, que é o que está descrito na mediana igual a zero. Mas em contraste, foi registrado situação em que os participantes apresentaram uma melhora de até 6 pontos, como mostra o valor descrito máximo da variável delta.

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
## -2.0000  0.0000  0.0000  0.5556  1.0000  6.0000      3
```

Analises da variação antes e depois

Analisando a variável delta, isto é, a melhora ou piora após o treinamento, temos o seguinte:

- 40.3% dos participantes melhoraram
- 10.5% não melhoraram, mas porque já otiveram nota máxima no teste inicial (portanto mantiveram a pontuação)
- 28.0% podiam melhorar mas continuar iguais
- 15.7% pioraram.

evolucao	status	n	(%)
Melhorou	Pode melhorar	23	40.35
Não alterou	Não tem como melhorar	6	10.53
Não alterou	Pode melhorar	16	28.07
Piorou	Pode melhorar	9	15.79
NA	Pode melhorar	3	5.26

Além disso, abaixo constam as frequências de notas antes e após o treinamento.

Pré treinamento

```
##
##  4  6  7  8  9 10
##  1  2 12 14 16 12
```

Pós treinamento

```
##
##  6  7  8  9 10
##  1  2 12 23 16
```

Análises das perguntas

Perguntas a serem respondidas:

- Qual é o percentual médio de acerto por questão, antes e depois da formação?
- Qual é o percentual médio de seleção por alternativa de cada questão, antes e depois da formação?
- Quais foram os principais movimentos de mudança de alternativa por questão (árvore de probabilidades)?

Análise das variáveis do conjunto de dados (2)

Summary do conjunto de dados

```
## N°.da.Seção      N°.da.P      Tipo.da.P      Título.da.P
## Mode:logical    Min.      : 1.0    Length:50      Length:50
## NA's:50         1st Qu.: 3.0    Class :character Class :character
##                 Median : 5.5    Mode  :character Mode  :character
##                 Mean   : 5.5
##                 3rd Qu.: 8.0
##                 Max.   :10.0

## Texto.da.P      Bônus.      Dificuldade    Resposta
## Length:50       Length:50      Min.      :1    Length:50
## Class :character Class :character 1st Qu.:1    Class :character
## Mode  :character Mode  :character Median :1     Mode  :character
##                                     Mean   :1
##                                     3rd Qu.:1
##                                     Max.   :1

## Relação.de.Respostas N°.de.Respostas Pontuação.Média      De
## Length:50           Min.      : 0.00    Length:50      Min.      :1
## Class :character    1st Qu.: 0.00    Class :character 1st Qu.:1
## Mode  :character    Median : 1.50    Mode  :character Median :1
##                                     Mean   :11.40    Mean   :1
##                                     3rd Qu.: 8.75    3rd Qu.:1
##                                     Max.   :56.00    Max.   :1

## Desvio.padrão      Índice.de.Discriminação Ponto.Bisserial
## Length:50          Length:50          Length:50
## Class :character    Class :character    Class :character
## Mode  :character    Mode  :character    Mode  :character
##
##
##

## N°.da.Seção      N°.da.P      Tipo.da.P      Título.da.P
## Mode:logical    Min.      : 1.0    Length:50      Length:50
## NA's:50         1st Qu.: 3.0    Class :character Class :character
##                 Median : 5.5    Mode  :character Mode  :character
##                 Mean   : 5.5
##                 3rd Qu.: 8.0
```

```

##           Max.      :10.0
##   Texto.da.P      Bônus.      Dificuldade   Resposta
## Length:50      Length:50      Min.      :1   Length:50
## Class :character Class :character 1st Qu.:1   Class :character
## Mode  :character Mode  :character Median :1   Mode  :character
##                                     Mean  :1
##                                     3rd Qu.:1
##                                     Max.   :1
## Relação.de.Respostas N°.de.Respostas Pontuação.Média      De
## Length:50      Min.      : 0.0   Length:50      Min.      :1
## Class :character 1st Qu.: 0.0   Class :character 1st Qu.:1
## Mode  :character Median : 1.0   Mode  :character Median :1
##                                     Mean  :10.8      Mean  :1
##                                     3rd Qu.: 6.0      3rd Qu.:1
##                                     Max.   :54.0      Max.   :1
## Desvio.padrão      Índice.de.Discriminação Ponto.Bisserial
## Length:50      Length:50      Length:50
## Class :character Class :character      Class :character
## Mode  :character Mode  :character      Mode  :character
##
##
##

```

Valores únicos em cada variável

Valores unicos em cada variavel, exceto variáveis 5 e 8 por se tratar do texto das questões e respectivas alternativas de “A” a “E”. A ideia geral com este procedimento é entender a varibilidade de cada variável para que seja possível direcionar o tipo de análise que será empreendida para responder as questões.

```

## $N°.da.Seção
## [1] NA
##
## $N°.da.P
## [1] 1 2 3 4 5 6 7 8 9 10
##
## $Tipo.da.P
## [1] "MC"
##
## $Título.da.P
## [1] ""
##
## $Bônus.
## [1] "Falso"
##
## $Dificuldade
## [1] 1
##

```

```
## $Relação.de.Respostas
## [1] "Marcado"      "Desmarcado"
##
## $Nº.de.Respostas
## [1] 56  0  1  2 54  8 43  4 46 12 42  9 30 17  5 50  7 49
##
## $Pontuação.Média
## [1] "0,98" "0,95" "0,75" "0,81" "0,74" "0,53" "0,88" "0,86"
##
## $De
## [1] 1
##
## $Desvio.padrão
## [1] "13,13%" "22,33%" "43,05%" "39,46%" "44,03%" "49,93%" "32,82%" "34,74%"
##
## $Índice.de.Discriminação
## [1] "6,67%" "13,33%" "53,33%" "60,00%" "26,67%" "73,33%" "20,00%"
##
## $Ponto.Bisserial
## [1] "0,46" "0,25" "0,51" "0,55" "0,26" "0,54" "0,35" "0,23" "0,27"
```

Formato das perguntas do teste

As questões em possuem a característica de multipla escolha, conforme a lista a seguir das três primeiras questões do teste.

Três primeiras perguntas:

- (1) Segundo princípios definidos pela LDB (art. 3º, inciso VIII) e pela Constituição da República art. 206, inciso VI), a educação é um processo social, construído por meio da participação da comunidade escolar. Uma boa prática de gestão democrática envolvendo a relação escola-família é:
 - (a) realizar atividades que envolvam a participação das famílias, compartilhando rotinas e decisões da escola.
 - (b) compreender a gestão democrática como um processo direcionado à equipe pedagógica da escola.
 - (c) centralizar as ações entre a escola e as famílias em uma liderança indicada pelos(as) próprios(as) estudantes entre os(as) seus(suas) responsáveis.
 - (d) permitir às famílias dos(as) estudantes realizar alterações no Projeto Político-Pedagógico da escola quando esses(as) se sentirem desconfortáveis com a gestão.
 - (e) delegar aos(às) vice-diretores(as) e coordenadores(as) pedagógicos(as) o controle da disciplina dos(as) estudantes da escola para que possam estabelecer comunicação com as famílias.
- (2) O diretor de uma unidade escolar se autodeclara praticante de uma gestão democrática. Essa declaração pode ser considerada verdadeira apenas se esse diretor convida a comunidade escolar para:
 - (a) coordenar o grêmio estudantil.

- (b) construir o Projeto Político-Pedagógico.
 - (c) elaborar o planejamento bimestral das disciplinas.
 - (d) executar o plano de curso de atividades extracurriculares.
 - (e) coordenar um programa de recomposição de aprendizagens.
- (3) Em regiões socialmente vulneráveis, nas quais existam mais de uma escola para atendimento à população local, uma boa prática de análise e resolução contextualizada de desafios que afetam internamente as instituições é:
- (a) a comparação dos resultados das escolas nas avaliações sistêmicas.
 - (b) o estabelecimento de metas educacionais iguais para as escolas locais.
 - (c) o acionamento das famílias responsáveis pelos(as) estudantes para cobrar providências.
 - (d) o diálogo entre os(as) diretores(as) escolares para compartilhamento de boas práticas, orientado pelas esferas superiores de gestão da Secretaria
 - (e) a intervenção das esferas superiores de gestão, que, no caso da educação brasileira, podem envolver a sede da Secretaria e regionais de ensino.

Seleção dos dados pré teste

No caso das variáveis perguntas e respostas trabalharemos com o número ao invés de texto. Recorreremos ao texto apenas quando necessário para analisar algum caso particular.

- **Qual é o percentual médio de acerto por questão, antes e depois da formação?**

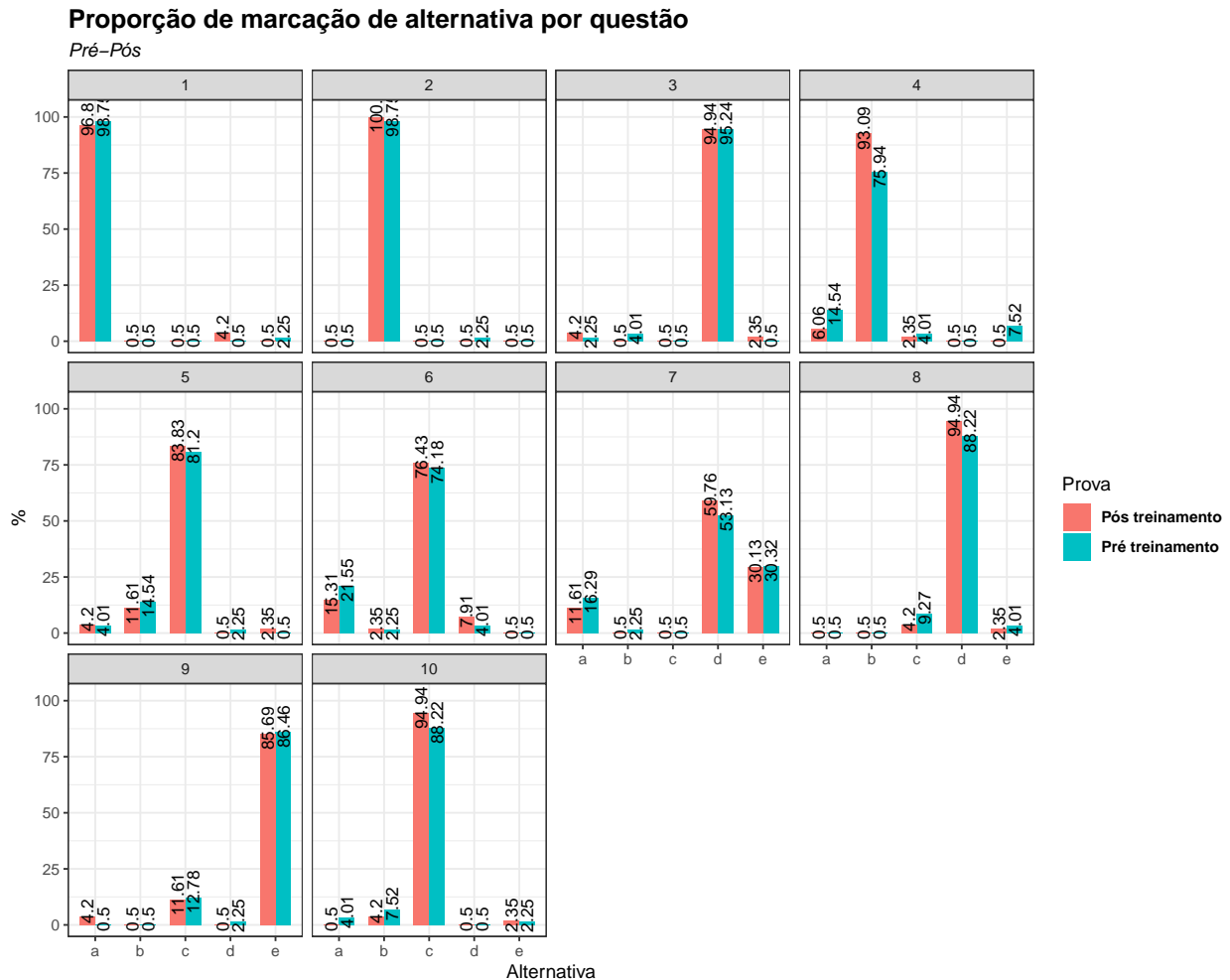
Note que 3 perguntas em particular tiveram variação negativa, isto é, em média, os participantes após o treinamento performaram pior nessa questão. Outro destaque nesta análise, diz respeito a pontuação média das questões (antes e depois) 6 e 7. Como apresentaram menor pontuação média, estas parecem indicar maior dificuldade, já que a pontuação média de ambas é menor e sua variação (delta) é marginal após o treinamento. Além disso o desvio padrão dela também sugere isso, já que ele indica alta variância, portanto maior dispersão (distância) em torno da média.

Nº da Pergunta	Pont. Média Pré	Pont. Média Pós	Delta	Desvio Pré	Desvio Pós
1	0.98	0.96	-0.02	13,13%	18,89%
2	0.98	1.00	0.02	13,13%	0,00%
3	0.95	0.94	-0.01	22,33%	22,91%
4	0.75	0.93	0.18	43,05%	26,19%
5	0.81	0.83	0.02	39,46%	37,27%
6	0.74	0.76	0.02	44,03%	42,75%
7	0.53	0.59	0.06	49,93%	49,14%
8	0.88	0.94	0.06	32,82%	22,91%
9	0.86	0.85	-0.01	34,74%	35,52%
10	0.88	0.94	0.06	32,82%	22,91%

As duas questões finais a seguir, se o entendimento estiver correto, podem ser respondidas com os gráficos que se seguem.

- Qual é o percentual médio de seleção por alternativa de cada questão, antes e depois da formação?

O que o gráfico a ilustra é a proporção de marcação de cada alternativa antes e após o treinamento. Ou seja, considere a questão 1 (primeiro gráfico da esquerda para direita) barra azul. Ele indica que antes do treinamento, 96.8% dos participantes marcaram a letra “a” como alternativa correta. Em contraste, a barra vermelha indica esse mesmo percentual após o treinamento. Note que as principais variações de marcação de alternativas ocorreram na questão número 4, alternativa “b”.



- Quais foram os principais movimentos de mudança de alternativa por questão?

Gráfico de variação por questão

Aqui o que gráfico evidência é o delta (%), isto é, a variação de marcação das alternativas antes e depois do treinamento. No gráfico 1, por exemplo, tem-se que o percentual de marcação total nas letras “a” e “e” caiu após o treinamento, ao passo que “b” e “c” não ocorreram variações e “d” uma variação positiva nas marcações.

Variação de marcação de alternativa por questão

Delta = (%) Pós treinamento – (%) Pré treinamento

